A DRAFT *DE NOVO* GENOME ASSEMBLY AND MITOCHONDRIAL POPULATION GENOMICS FOR THE NORTHERN BOBWHITE (*COLINUS VIRGINIANUS*)

A Dissertation

by

YVETTE ASHTON HALLEY-SCHULTZ

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Christopher M. Seabury
Committee Members,	William J. Murphy
	Michael F. Criscitiello
	James E. Womack
Intercollegiate Faculty Chair,	Craig J. Coates

December 2015

Major Subject: Genetics

ABSTRACT

Wild populations of northern bobwhites (*Colinus virginianus*; hereafter bobwhite) have declined across most of their historic U.S. range, and despite their importance as an experimental wildlife model for ecotoxicology studies, no bobwhite draft genome assembly has emerged. Herein, we present the first bobwhite draft *de novo* genome assembly, with more than 90% of the assembled bobwhite genome captured within < 40,000 final scaffolds (N50 = 45.4 Kb) despite evidence for approximately 3.22 heterozygous polymorphisms per Kb. Moreover, three annotation analyses produced evidence for > 14,000 unique genes and proteins. Bobwhite analyses of divergence with the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*) genomes revealed many extremely conserved gene sequences, and evidence for lineage-specific divergence of noncoding regions. Coalescent models for reconstructing the demographic history of the bobwhite and the scarlet macaw were concordant with how opposing natural selection strategies (i.e., skewness in the *r*-/*K*-selection continuum) would be expected to shape genome diversity and the effective population sizes in these species.

Using genomic tools and resources developed via the draft genome assembly, we evaluated the concordance of population inferences and conclusions resulting from the analysis of short mitochondrial fragments (i.e., partial or complete D-Loop nucleotide sequences) versus complete mitogenome sequences for 53 bobwhites representing six ecoregions across TX and OK (USA). Median joining (MJ) haplotype networks demonstrated that analyses performed using small mitochondrial fragments were

insufficient for estimating the true (i.e., complete) mitogenome haplotype structure, corresponding levels of divergence, and maternal population history of our samples. Notably, discordant demographic inferences were observed when mismatch distributions of partial (i.e., partial D-Loop) versus complete mitogenome sequences were compared, with the reduction in mitochondrial genomic information content observed to encourage spurious inferences in our samples. A probabilistic approach to variant prediction for the complete bobwhite mitogenomes revealed 344 segregating sites corresponding to 347 total mutations, including 49 putative nonsynonymous single nucleotide variants (SNVs) distributed across 12 protein coding genes. Evidence of gross heteroplasmy was observed for 13 bobwhites, with 10 of the 13 heteroplasmies involving one moderate to high frequency SNV. Haplotype network and phylogenetic analyses for the complete bobwhite mitogenome sequences revealed two divergent maternal lineages (d_{XY} = 0.00731; F_{ST} = 0.849; $P < 0.05$), thereby supporting the potential for two putative subspecies. However, the diverged lineage (n = 103 variants) almost exclusively involved bobwhites geographically classified as *Colinus virginianus texanus*, which is discordant with the expectations of previous geographic subspecies designations. Tests of adaptive evolution for functional divergence (MKT), frequency distribution tests (D, F_S) and phylogenetic analyses (RAxML) provide no evidence for positive selection or hybridization with the sympatric scaled quail (*Callipepla squamata*) as being explanatory factors for the two bobwhite maternal lineages observed. Instead, our analyses support the supposition that two diverged maternal lineages have survived from pre-expansion to post-expansion population(s), with the segregation of some slightly deleterious nonsynonymous mutations.

DEDICATION

I would like to dedicate this dissertation to my family. To my parents, Wendell and Monica Halley, and my little brother, William Halley, for all of their love, friendship, support and sacrifices; I am truly blessed to have belonged to such an incredible family. I would also like to dedicate this dissertation to my husband, Kyle Schultz. Thank you for being so understanding, loving and patient, I couldn't have asked for a more supportive partner in life. Lastly, I would like to acknowledge all Texas A&M University former students who have donated to the Student Assistance fund, the assistance I received has left me eternally grateful.

ACKNOWLEDGMENTS

Foremost, I am very grateful to my committee chair, Dr. Christopher M. Seabury, for all of his support, guidance and encouragement throughout my graduate education. I would also like to thank Dr. William J. Murphy, Dr. James E. Womack, and Dr. Michael F. Criscitiello for serving on my committee and providing me with valuable advice. I also acknowledge the help and support provided by Eric K. Bhattarai and David L. Oldeschulte; thank you for your comradery and friendship. Additionally, I would also like to extend my thanks to all my collaborators whose cooperative efforts helped make the work within this dissertation possible: Dr. Jared Decker, Dr. Scot Dowd, Dr. Joshua Hill, Dr. Charles Johnson, Dr. Richard Metz, Dr. Markus J. Peterson, Dr. Steven Presley, Dr. Dale Rollins, Rebekah Ruzicka, and Dr. Jeremy F. Taylor.

This project wouldn't have been possible without the financial support of Joe Crafton and Park Cities Quail, the Rolling Plains Quail Research Foundation, and the Reversing the Quail Decline in Texas Initiative with Upland Game Bird Stamp Fund, based on a collaborative effort between Texas Parks and Wildlife Department and Texas A&M AgriLife Extension Service. Thank you to all the hunters who donated samples, and to Texas Parks and Wildlife Department, I am also grateful to Doug Schoeling, Jena Donnell, and the Oklahoma Department of Wildlife. Lastly, I would like to extend my thanks to the Texas AgriLife Genomics and Bioinformatics Core, Texas A&M University, and the Missouri Sequencing Core (Nathan Bivens; Sean Blake) at the University of Missouri for high quality sequencing services.

NOMENCLATURE

Abbreviation	Description
cDNA	Complementary DNA
BAC	Bacterial Artificial Chromosome
PE	Paired-end
MP	Mate pair
Gbp	Gigabase Pairs
Kbp	Kilobase Pairs
SNP	Single Nucleotide Polymorphism
MYA	Million Years Ago
ORF	Open Reading Frame
MHC	Major Histocompatibility Complex
GWAS	Genome-wide Association Studies
SAM	Sequence Alignment/Map Format
SRA	Sequence Read Archive
WGS	Whole Genome Shotgun
LINE	Long Interspersed Elements
SINE	Short Interspersed Elements
LTR	Long Terminal Repeats
KYA	Thousand Years Ago
PSMC	Pairwise Sequentially Markovian Coalescent

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Bobwhite Biology and Ecology

The northern bobwhite (*Colinus virginianus*; hereafter bobwhite) was named for its distinct mating call (i.e., ah-bob-white) (1-4). Historically their native range included savannas, open woodlands, and brushy grasslands throughout the United States of America (USA), Mexico and parts of the Caribbean. The bobwhite is one of thirty-two species belonging to the family Odontophoridae (New World Quail) (5-7). Within this family, the bobwhite is perhaps the most diverse, with nineteen named subspecies varying both in size (i.e., increasing from south to north; similar to Bergmann's rule) (8) and morphology (5). Specifically, the most overt morphological variation occurs on the head and underparts, which are marked by variable combinations of grey, brown, and white, with males showing the highest level of plumage color variation across their geographic range (5, 9, 10). Wild bobwhite populations in the southern U.S. and northern Mexico have been divided into four putative subspecies that occur west of the Mississippi River, and include the eastern (*C. v. virginianus*), plains (*C. v. taylori*), Texas (*C. v. texanus*), and masked bobwhite (*C. v. ridgwayi*) (9, 10). Among these, the male masked bobwhite is unequivocally the most phenotypically and geographically distinct (i.e., black head; Sonora, Mexico), but at present, is also the only endangered quail in North America (11). In comparison, the eastern, plains, and Texas bobwhites exhibit more subtle variation in male plumage and body size (9, 10). Among the different subspecies of bobwhites,

females are virtually indistinguishable from one other, with males and females exhibiting pronounced sexual dimorphism (6, 11).

The bobwhite is one of the most voca1ly communicative species of the order of Galliformes, as evidence by more than twenty documented calls (1, 12) that have been interpreted to include a number of different meanings (i.e., group movement, food finding, predator avoidance, reproduction, social) (1). Solitary birds have social calls that can be classified into two different groups: bringing males and females together, or spacing out males (1). The bobwhite is a gregarious bird, and therefore, requires the use of a third social call: regrouping scattered members of a group (1). Generally considered a non-migratory terrestrial bird, the bobwhite has been considered to possess low mobility (13, 14). The habitat characteristics required by the bobwhite includes plant communities that provide important variables such as nesting and brooding cover, thermal cover, predation cover, and food (7, 13, 15-20). Bobwhite roosting sites are generally categorized by 'stopping points' (i.e., nocturnal roosts, mid-day cover) (20). The bobwhite social system consists of a combination of both solitary and group living environments (i.e., pair, brood, coveys) (2, 14, 21-23). Social group formation is hypothesized to be dependent on the advantages of thermoregulation (14, 24), predatory evasion (14, 25, 26), and foraging efficiency (14, 24, 26). Bobwhites typically roost in a circle, such that there is always a bird that can discover danger (27). Natural (i.e., coyote, owls, hawks, foxes, weasels, raccoons, bobcats) and introduced (i.e., house cats) predators are of concern for bobwhites (28, 29). However, bobwhites are usually not a main food source for their predators (30), and the bobwhite's risk of predation is hypothesized to be largely a function of the

behavior and densities of predators in relation to the behavior and densities of alternate prey species (31).

The bobwhite's breeding season typically spans from late spring till early fall (32-34); with sexual behavior usually limited to this period, and both males and females exhibiting their own mannerisms (1). The male's sexual behaviors are classified as: lateral display, bowing, tidbitting, food calls, copulation, nesting ceremony, and bobwhite calls (1). The female's sexual behavior consists of nudging, wing quivering, presentation, precopulatory crouch, copulation with associated call, and nest ceremony (1). Females usually are able to lay their first eggs when they are approximately 20 - 22 weeks of age (4). Interestingly, bobwhites are generally considered monogamous breeders with singular clutches (15, 35-38), but have been known to produce rapid multi-clutches, and may exhibit ambisexual polygamous mating systems (2, 39-43). Bobwhites are ground nesting birds (1, 27, 44), with nesting success greatly influenced by weather (i.e., rainfall and temperature) (2, 37, 45). Nests are typically found in meadows, green fields or along fence rows and roadsides that are grown up with grass or weeds; built of arched-over grasses such that nest appears like a tunnel, which is ideal for hiding eggs (27). Natural predation is known to be a major cause of nest failure (15, 39, 46-49); with predators ranging from mammas (i.e., foxes, weasels, opossums, coyotes, raccoons, bobcats, rodents) and reptiles (i.e., snakes), to fire ants (*Solenopsis invicta*) (15, 37, 46, 48, 50, 51). If a nest is destroyed the breeding pair will customarily work rapidly to get their next nest constructed (27, 39). Both sexes usually participate in nest building and incubation (1, 44), with males incubating on average 13-27% of the nests (15, 37, 39, 41, 52). In the multiclutch mating

system, females lay a clutch that is incubated by a male, and then the female lays and incubates another clutch (39). Egg incubation of the bobwhite is ~ 24 days (53, 54), and the mean clutch size is 15.6 eggs for the first clutch and 12.8 eggs for the second clutch (37). Double clutching is another nesting habit exhibited by bobwhites. In double clutching, a female will produce and hatch an initial clutch and then attempt to lay a second nest (39). In general, roughly 25% or more of bobwhite females will attempt double clutching during a breeding season (39, 41, 52, 55).

Relevant to reproduction, bobwhites exhibit relatively overt characteristics of an *r*-selected species (i.e., produce many offspring that have a low probability of survival, early maturity), and produce numerous precocial (i.e., chicks can fend for themselves and are self-thermoregulatory) young (2, 14, 56-60), with chick mortality being highest during the first few weeks of life (59, 61, 62). Bobwhite chicks are very small, downy, and brown-streaked; with this color pattern rendering them nearly invisible when motionless (27). For bobwhite chicks to be successful, a high protein diet is required, which is typically provided by invertebrates (i.e., grasshoppers, chinch bugs, mosquitos, plant lice, mealworms, crickets, etc.) (27, 59, 61-64); commonly eating their weight in insects daily (27). Juvenile bobwhites graduate to adult eating habits around 7 – 9 weeks of age (59), where their adult diet is comprised of a variety seeds (i.e., crab grass, cockspur, foxtail, rib grass, pig weed, chickweed, acorn, beechnuts, chestnuts, pine), fruits (i.e., mulberry, blackberry, raspberries, sumac, grapes, sour gum, honeysuckle), plants (i.e., sorrel, cinquefoil, clover), grains, and animal foods (i.e., invertebrates) (27, 59, 64); with the dietary needs of the two sexes (i.e., protein intake) interrelated with the season (16, 59, 65,

66). As demonstrated in 'crop' studies, females typically consumed more animal food and more food overall than males during breeding seasons (59), which is unsurprising given the expected nutritional requirements for egg production.

Bobwhite Population Decline in the U.S.

Historically, the relative abundance of bobwhites across their native range has often been described as following a boom-bust pattern, with substantial variation in abundance among years (44, 67-69). Although broad scale declines in bobwhite abundance probably began somewhere between 1875 and 1905 (70-72), several better quantified studies of this long-term decline utilizing either breeding bird surveys or Christmas bird count data were reported beginning more than 20 years ago (73-78). This range-wide decline in bobwhite abundance across most of the U.S. is still ongoing today (79, 80), with bobwhite coveys in the southeastern portion of their range declining the most rapidly (73, 76, 81).

It has been exceedingly challenging to identify specific sources of the range-wide declines in the U.S., and in particular, to attribute this population trend to any individual factor. Therefore, the precise reasons for recent population declines in the U.S. appears to be complex, and have been attributed to many factors: variation in annual rainfall (67-69), thermal tolerances of developing embryos within a period of global warming (82, 83), shifts in land use and scale coupled with the decline of suitable habitat (44, 67, 77, 78, 84), red imported fire ants (i.e., predation, hazardous food source) (85-87), sensitivity to ecotoxins (88, 89), harvest intensity by humans (90-92), particularly during drought conditions (69, 84), and most recently, parasitic eyeworms (93-95). However, it should

also be noted that the same species of parasitic eyeworm (*Oxyspirura petrowi*) detected in wild bobwhites have also been detected in songbird species (i.e., Northern Mockingbird, *Mimus polyglottos)* and Curve-billed Thrasher (*Toxostoma curvirostre*) (96) which have not been reported to be experiencing similar broad-scale population declines (http://www.iucnredlist.org).

Population declines have prompted efforts to translocate bobwhites to fragmented parts of their historic range where modern abundance is low. However, the results of these translocations have proven to be highly variable (97-99), with one such recent study demonstrating that bobwhites fail to thrive in historically suitable habitats that have since become fragmented (99). Fragmentation of historic bobwhite habitats increases the possibility of inbreeding within isolated populations (i.e., pockets) due to reduced gene flow between discrete populations (6, 14). Restocking via the release of pen-reared bobwhites has also been explored, with all such efforts reporting low survival rates (97, 100-102), and those that do survive may potentially dilute local genetic adaptations via successful mating with remnant members of wild populations (102).

Bobwhite Harvest in the U.S.

In 1883, Alfred Mayer said, "Of all the game birds of America, none is better appreciated by the sportsman than little Bob White" (103). Bobwhite hunting is a U.S. tradition, rooted and steeped in southern culture (104, 105), with the bobwhite arguably being one of America's most economically important (6, 104, 106) and popular gamebirds (27, 103). In the State of Texas, the bobwhite hunting season generally spans for

approximately 120 days, with daily 'bag' limits of 15 birds per hunter (107). Bobwhite hunting affects both non-hunting and hunting related industries; with economic contributions to both state and federal agencies (104). In 1991 it was predicted that huntable bobwhite populations would vanish by 2005 (73); with the observed decline between 1991 and 1992 estimated to have caused a loss of 13.3 million dollar loss to Southeastern U.S. rural communities via decreased hunting expenditures (42, 104). Bobwhite populations in Texas have declined at an estimated rate of 5.6% per year since 1980 (Texas Parks and Wildlife 2005) (106). This decline has had an adverse effect on the hunting community and related economy, such as those in southeast Texas, where historically wildlife populations were a by-product of agriculture (108). The prospect of economic loss has created a need for management activities and programs that would increase wild huntable populations, and decrease hunter attrition (104, 105). A decrease in 'wild' hunting opportunities, declining harvest numbers, and the noted decline in hunter numbers has led to alternative management practices, such as the release of pen-reared bobwhites to supplement hunting needs (102). Bobwhite quail hunting brings in an estimated state sales tax of ~$4.5 million, state income tax revenue of ~$1.8 million, and federal income tax revenue of ~$6.8 million from the southeastern U.S. (i.e., AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA); although these estimates are conservative since true regional impacts would account for interstate trade (104). Moreover, in 2006, it was reported by the Congressional Sportsmen's Foundation that Texas led the nation in hunter-related expenditures ($2.3 billion) (108), with a tangible proportion of those expenditures related to quail hunting (i.e., bobwhite and scaled quail; *Callipepla squamata*). For these

reasons, it is currently not uncommon for ranchers to prioritize deer and bobwhite conservation over cattle ranching (108).

In the 1950's Herbert Stoddard predicted that bobwhite hunting would become an expensive and rare experience that would be available only to the wealthy (109). Historical documents note that in the early 20th century, bobwhites were abundant and provided many easily accessible hunting opportunities throughout the Midwest and Southeast (70, 109, 110). Historically, people of moderate means were able to hunt bobwhites (109). However, at present, most quail hunting is only accessible to people who have the financial means to own or lease vast stretches of land, and who are also financially able to absorb land management costs ($50 - $200/ha/year) (109).

Historic Bobwhite Research

In 1905, Sylvester Judd published a book entitled, *The Bobwhite and Other Quails of the United States in their economic relations*. In his book, Judd encouraged sportsmen, farmers, legislators and ornithologists to take an interest in bobwhite preservation by citing how their food habits (i.e., destroying weed seeds, consuming pests) are economically important to farmers, the economic impact of hunting, their potential as a food source, and the aesthetic pleasure of their presence (3). Historically, bobwhite population studies were the fundamental building blocks in the early development of modern aspects of wildlife ecology (15, 111-114), with organized research focusing on the bobwhite tracing back to the 1920's (15, 115, 116). To date, very few wildlife species have received the same levels of research and management attention (117). Over the past century, many bobwhite

studies have focused on species biology, ecology and management (6, 109, 118). However, at the start of the 20th century a majority of bobwhite publications focusing on their geographic distribution, general importance as a gamebird, and the start of their decline (3, 119-123). With some of the earliest recorded publications documenting the feeding habits of the bobwhite (3, 65). By the 1940's bobwhite research had graduated to ecological and behavioral studies; with a few studies centering around courtship and behavior (124, 125). During the period that these studies took place, it was noted that bobwhites engage in courtship feeding, and the function of this behavior most likely to maintain a mating bond (124). Studies also arose which proposed more detailed protocols on how to classify bobwhites by age (15, 126). Behavioral studies recognized that adult and juvenile mannerisms differ, and to properly classify behavior created a useful tool for enabling biologists to separate the age-classes suitably (126). Thereafter, Stoddard and Errington were among the first to suggest that bobwhites were sedentary birds who spent most of their lifetimes within 400 meters of their hatching site (15, 115). However skepticism led to more studies that challenged this notion (127), indicating that bobwhites were capable of long-distance travel (128).

Historical bobwhite ecology studies have covered a vast array of topics and hypotheses. For example, nesting studies provided a foundation and general understanding of bobwhite nesting behavior (129), aspects of nest predation (129), and egg incubation (130). Vocalization studies were also an important focus of early bobwhite density studies, and were employed by counting vocalizations on a represented sample area and were then extrapolated to a larger area to give a numerical measure of population status (131).

Summer whistling cock counts proved useful for helping to predict fall and winter bobwhite population densities and hunter success (i.e., the average number of bobwhites bagged per gun hour) the following fall because a male making a "bob-white" call gave insight into the size of the breeding population (132-134). Morning calls (i.e., the number of coveys giving the assembly call) were historically utilized to index population density (7, 135) and to determine bobwhite response to brush management patterns (136). Historic life history studies also led to a better (i.e., modern) understanding of brood ecology (127). Collectively, these studies were able to create a foundation for quantifying brood and chick survival (17, 137, 138).

By the later part of the 20th century, the primary focus of bobwhite literature had shifted to aspects of applied bobwhite management (127, 139, 140). These management studies investigated habitat use by bobwhites and the habitat requirements needed to support a population (7, 18, 117, 140-143). During this time, intense predator control was also explored, with targeted predator removal providing limited degree of enhanced reproductive success among bobwhites (144, 145). Management studies later progressed to population-level and broad-scale demographic issues for the bobwhite, and have remained a major focus for assessing clues on managing the bobwhite decline (127). These studies reformed the traditional species-habitat models that previously only factored in biotic and abiotic habitat features to now include other variables (i.e., weather, population history), and demonstrated how to use these new models to implicate a better suited management strategy (76, 97, 99, 111, 117, 127, 143, 146).

Bobwhite Research Needs

At present, the bobwhite is one of the most broadly researched and intensively managed wildlife species in North America (67, 84, 147). The suitability of the bobwhite as a model wildlife species (i.e., sentinel species) for climate change, land use, toxicology, and conservation studies has also been well established (67, 73, 74, 84, 147-152). Despite all the intensive research attention, the bobwhite remained without an annotated draft genome assembly, thereby precluding genome-wide studies of extant wild bobwhite populations, and the utilization of this information to positively augment available management strategies. Likewise, utilization of the bobwhite as an experimental wildlife model cannot be fully enabled in the absence of modern genomic tools and resources.

Historically, little genome-wide sequence and polymorphism data have been reported for many wildlife species, thereby limiting the implementation of genomic approaches for addressing key biological questions for these species (153, 154). Cytogenetic analyses have demonstrated that the bobwhite diploid chromosome number is 2n = 82, which includes 5 pairs of autosomal macrochromosomes and the sex chromosomes, 8 pairs of intermediately sized autosomes, and 27 pairs of autosomal microchromosomes (155, 156). Recent bobwhite genomic efforts were performed by Arun Rawat and colleagues, who utilized pyrosequencing of a normalized bobwhite multi-tissue cDNA library (assembled from both sexes) to construct an *in silico* annotation to characterize the transcriptome and create a custom microarray that could be utilized to explore the molecular impacts of exposure to 2,6-dinitrotoluene (2,6-DNT) (152) and to perform comparative studies with the annotated domestic chicken (*Gallus gallus*) genome

(157). Yet, no genome maps (i.e., linkage, radiation hybrid, BAC tiling paths) exist for the bobwhite.

More recently, a bobwhite mitochondrial DNA (mtDNA) study for the four putative subspecies described west of the Mississippi River was conducted (9), and focused exclusively on a 353 bp fragment corresponding to the bobwhite mitochondrial control region. This work reported a general lack of distinct phylogeographic structure, evidence for demographic expansion following the Pleistocene (i.e., unimodal mismatch distribution), and an apparent discordance between patterns of mtDNA diversity and geographic subspecies designations (9). Nonetheless, a need currently exists to establish genomic tools and resources that will facilitate large-scale genetic studies in the bobwhite, especially considering the apparent decline of wild bobwhite populations across the majority of their historic U.S. range (73, 74, 77-80). Herein, we developed and used such resources to draw inferences about historic and modern aspects of bobwhite biology.

CHAPTER II

GENOME SEQUENCING AND *DE NOVO* ASSEMBLY[*]

Introduction

Since the introduction of the chain termination sequence method (i.e., Sanger sequencing) in 1977 (158) and second-generation sequencing (SGS) in 2005 (159), DNA sequence technologies have literally shaped a biological sciences revolution (160). To date the genomes of more than 800 bacteria and 100 eukaryotes have been assembled (160). Moreover, the emergence of next generation sequencing (NGS) technologies, with their reduced generation time, enhanced yield, and low cost (159), in conjunction with enhanced bioinformatics tools, have essentially catalyzed a ‘‘genomics-era’’ for underrepresented model species (154). Evidence for this “genomic-era” has become particularly increasingly apparent with the many new avian genome sequence assemblies that have emerged for species like the Puerto Rican parrot (*Amazona vittata*) (161), the flycatchers (*Ficedula spp*) (162), the budgerigar (*Melopsittacus undulatus*; http://aviangenomes.org/budgerigar-rawreads/), the saker and peregrine falcons (*Falco peregrinus*; *Falco cherrug*) (163), Darwin’s finch (*Geospiza fortis*; http://gigadb.org/darwinsfinch/), the scarlet macaw (*Ara macao*) (154), the pigeon (*Columba livia*) (164), the black grouse (Tetrao tetrix) (165), the orange-bellied parrot (*Neophema chrysogaster*) (166), and the ground tit (*Pseudopodoces humilis*) (167).

[*] Portions of this chapter were reprinted with permission from “A Draft *De Novo* Genome Assembly for the Northern Bobwhite (*Colinus virginianus*) Reveals Evidence for a Rapid Decline in Effective Population Size Beginning in the Late Pleistocene” by Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, Rollins D, Tizard IR, Brightsmith DJ, Peterson MJ, Taylor JF, Seabury CM, 2014. PLoS ONE 9(3): e90240. doi:10.1371/journal.pone.0090240, Copyright 2014 Halley et al.

Relevant to the bobwhite, recent genomic efforts have focused on generating cDNA sequences for the construction of a custom microarray (8,454 genes) to study the physiological effects of ecotoxicity (152), and for comparative studies with the annotated domestic chicken (*Gallus gallus*) genome (157). However, no genome maps (i.e., linkage, radiation hybrid, BAC tiling paths) exist for the bobwhite. For this present study, a single wild female (i.e., Pattie Marie) was utilized to construct the first bobwhite draft *de novo* genome assembly which consisted of, both simple contigs and a scaffolded *de novo* assembly, with each assembly utilizing greater than 2.3 billion next generation sequence reads produced from paired-end (PE) and mate pair (MP) Illumina libraries.

Results and Discussion

Herein, we assembled a genome sequence for Pattie Marie, a wild, adult female bobwhite from Texas. All sequence data were generated with the Illumina HiSeq 2000 sequencing system (v2 Chemistry; Illumina Inc.; San Diego, CA). As previously described (154), we estimated the bobwhite nuclear genome size to be ≈ 1.19–1.20 Gigabase pairs (Gbp; See Methods). While this estimate does not fully account for the lack of completeness in all existing avian genome assemblies (i.e., collapsed repeats), it is useful for determining whether the majority of the bobwhite genome was captured by our *de novo* assembly. Collectively, more than 2.36 billion trimmed sequence reads derived from three libraries (see Methods) were used in the assembly process (Table 1), which yielded ≥ 142X theoretical genome coverage (1.19–1.20 Gbp) as input data, and ≥ 77X assembled coverage (Table 2). Summary and comparative data for major characteristics of the

bobwhite draft *de novo* genome assembly are presented in Table 2, which also includes a comparison to the initial releases of two established and well annotated avian reference genomes from the order Galliformes (168, 169).

To assess the consistency of our assembly and scaffolding procedures, and to facilitate fine-scale analyses of divergence as previously described, we produced a simple *de novo* (i.e. no scaffolding; hereafter NB1.0) and a scaffolded *de novo* assembly (hereafter NB1.1), with the scaffolding procedure using both PE and MP reads to close gaps and join contigs. The concordance between the two assemblies was profound, with > 90% of the simple *de novo* contig sequences mapping onto the scaffolded assembly with zero alignment gaps (Table 2, Table S1) (170). Our first generation scaffolded assembly contained 1.172 Gbp (including N's representing gaps; 1.047 Gbp of unambiguous sequence) distributed across 220,307 scaffolds, with a N50 contig size of 45.4 Kbp (Table 2). Moreover, > 90% of the assembled genome was captured within < 40,000 scaffolds (Fig. 1). Importantly, these results meet or exceed similar quality benchmarks and summary statistics initially described for several other avian genome assemblies (i.e., Puerto Rican parrot, scarlet macaw, chicken, turkey) (154, 161, 168, 169), but do not exceed summary statistics (i.e., scaffold N50, etc.) for some recent assemblies (i.e., Flycatcher, Peregrine and Saker Falcons) that utilize either ultra-large insert mate pair libraries and/or available maps for enhanced scaffolding (162, 163).

Table 1. Summary of Illumina Sequence Data Used for *De Novo* Genome Assembly.

Data Source	Total Reads[a]	Library Type	Insert Size PD Dist. (bp)[b]	Average Read Length (bp)[c]
Illumina HiSeq	1,575,625,135	Small Insert Paired End	230-475[c]	84
Illumina HiSeq	510,031,444	Mate Pair (Small)	2100-3100[c]	49
Illumina HiSeq	276,134,302	Mate Pair (Medium)	4600-6000[c]	50

[a] Total usable reads after quality and adapter trimming (n = 2,361,790,881).
[b] Insert size and corresponding range of paired distances for each Illumina sequencing library.
[c] Averages for quality and adapter trimmed reads, rounded to the nearest bp.

Table 2. Comparison of Northern Bobwhite *De Novo* Genome Assembly to the Initial Turkey and Chicken Genome Assemblies.

Genome Characteristics	Simple *de novo* Bobwhite 1.0[a]	Scaffolded Bobwhite 1.1[b]	Turkey 2.01	Chicken 1.0
Total Contig Length[c]	1.042 Gbp	1.047 Gbp	0.931 Gbp	1.047 Gbp
Total Contigs > 1Kb	198,672	65,833	128,271	98,612
N50 Contig Size	6,260 bp	45,400 bp	12,594 bp	36,000 bp
Largest Contig	163,812 bp	600,691 bp	90,000 bp	442,000 bp
Total Contigs	374,224	220,307	152,641	NA[d]
Contig Coverage	≥ 100x[e]	≥ 77x[f]	17x	7x
Cost (M = million)	< $0.020M[g]	< $0.020M[g]	< $0.250M	> $10M

[a] No scaffolding procedure implemented (NB1.0).
[b] Scaffolding based on paired reads (NB1.1); no genome maps or BACs were available.
[c] Excluding gaps; scaffolded assembly with gaps (i.e., N's) = 1.172 Gbp.
[d] Not provided; see 46.
[e] Median and average coverage, excluding contigs with coverage > 300X (n = 4,293).
[f] Median and average coverage, excluding contigs with coverage > 300X (n = 3,717).
[g] The one-time cost of sequencing also reflects all library costs.

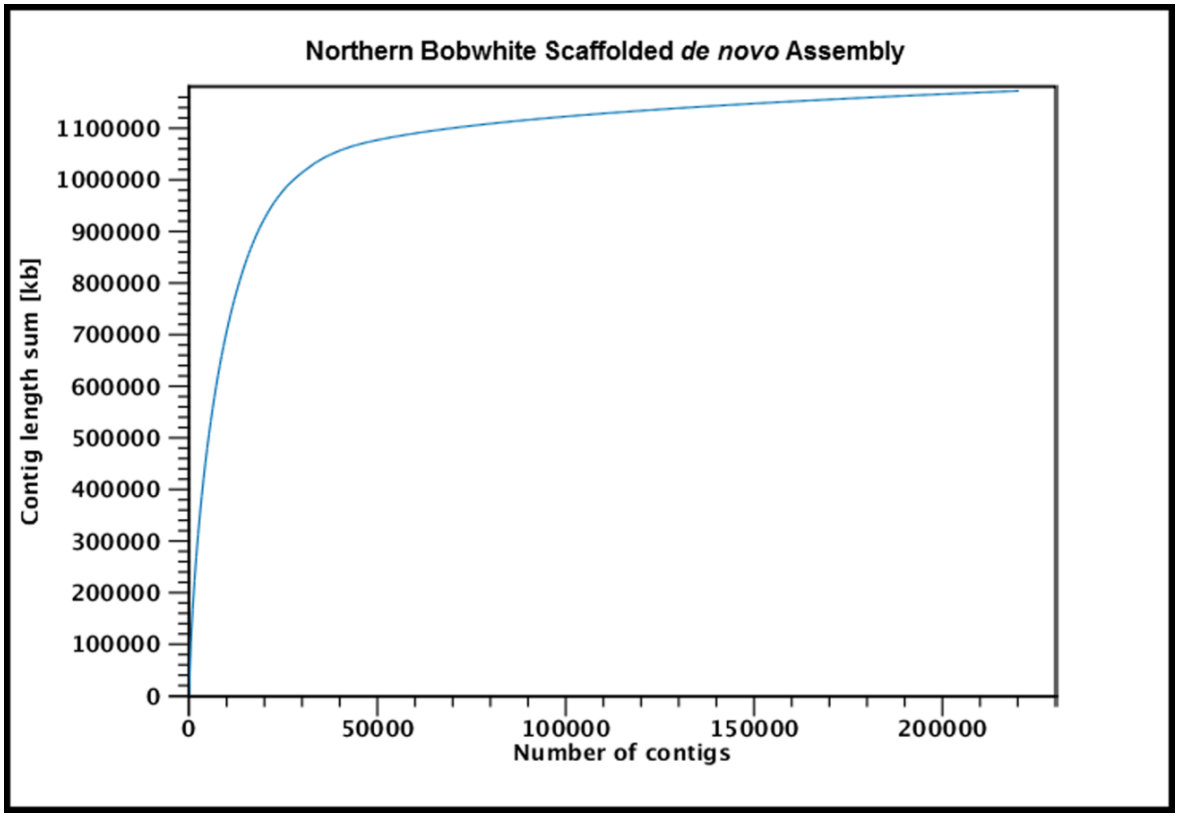


Figure 1. Relationship between Total Contig Length (Kbp) and Total Contig Number for the Scaffolded Bobwhite (*Colinus virginianus*) Genome (NB1.1). The y-axis represents total contig length, expressed in kilobase pairs (Kbp), and the x-axis represents the total number of scaffolds. The bobwhite genome was estimated to be 1.19-1.20 Gbp. For NB1.1 (1.172 Gbp), > 90% of the assembled genome was captured within < 40,000 scaffolds.

Methods

*Source of Bobwhite (*Colinus virginianus*) Genomic DNA*

We utilized skeletal muscle derived from the legs of a wild, female bobwhite ("Pattie Marie") from Fisher county Texas to isolate high molecular weight genomic DNA using the MasterPure DNA Purification Kit (Epicentre Biotechnologies, Inc., Madison, WI). Ethical clearance is not applicable to samples obtained from lawfully harvested wild bobwhites. The protocol for isolating genomic DNA followed the manufacturer's recommendations, and we confirmed the presence of high molecular weight genomic DNA by agarose gel electrophoresis, with subsequent initial quantification of multiple individual isolates performed using a Nano Drop 1000 (Thermo Fisher Scientific, Wilmington, DE).

Genome Sequencing Strategy

Prior to library construction, bobwhite genomic DNA was quantitated using the Qubit DNA HS assay and Qubit 2.0 flourometer (Life Technologies Inc., Carlsbad, CA), with further evaluation by agarose gel electrophoresis. All samples contained high molecular weight DNA >15 kb, with little or no degradation, thereby making them suitable for PE and MP library preparation. For creation of a small insert PE library, approximately 1.0 µg of DNA was normalized to 40 µl and fragmented to approximately 300 bp using the QSonica plate sonication system (Qsonica Inc., Newton CT). The fragmented DNA was blunt-end repaired, 3′ adenylated and ligated with multiplex compatible adapters using the NEXTflex DNA Sequencing Kit for Illumina (Bioo Scientific cat # 514104) prior to size

selection (200–400 bp fragments) using SPRI beads (Agencourt Inc., Brea CA). PCR enrichment was performed to selectively amplify bobwhite DNA fragments with adapters on both ends as follows: 98°C for 30 sec, 10cycles [98°C for 10 sec, 65°C for 30 sec, 72°C for 60 sec], 72°C for 5 minutes, 10°C hold. Bobwhite PE library validation was performed using the Bioanalyzer 2100 High Sensitivity DNA assay (Agilent Inc., Santa Clara, CA), with quantitation performed using the Qubit HS DNA assay. Thereafter, two MP sequencing libraries (Table 1) were created by following the Illumina Mate Pair v2 Library Preparation procedure for 2–5 Kbp fragments (Part #15008135 Rev A; Illumina Inc., San Diego, CA) as recently described [42]. The final PE and MP libraries were diluted to 10 nM in preparation for sequencing on a HiSeq 2000 genetic analysis system (Illumina Inc., San Diego, CA). The bobwhite PE library was processed using PE-100 cycle runs (2×100 bp), and the MP libraries were processed using MP-50 cycle runs (2×50 bp), with data generation (i.e., image processing and base calling) occurring in real time on the instrument. All clustering and base-calling was performed as recommended by the manufacturer. A summary of Illumina reads for all libraries is provided in Table 1. Prior to assembly, we used knowledge of avian genome size (nuclear DNA content, C-value) (171) in conjunction with physical knowledge of modern avian genome assemblies (bp) to estimate the size of the bobwhite nuclear genome (154).

Genome Assembly

Prior to assembly, all Illumina sequence reads were first trimmed for quality and adapter sequences using the CLC Genomics Workbench. Briefly, Phred quality base scores

(Q) were converted into error probabilities, read-based running sums for quality were calculated, and reads were trimmed as recently described (154). Following initial quality trimming, a second algorithm was used to trim ambiguous nucleotides (N) from the ends of every sequence read by referring to a user-specified maximum number of ambiguous nucleotides allowed (n = 2) at each end of the sequence, with subsequent removal of all other ambiguous bases. Finally, we also used the Workbench (i.e. Smith-Waterman algorithm) to specify, identify, and remove all sequencing adapters that could potentially be present in our sequence reads.

For the simple *de novo* (NB1.0) and the scaffolded assemblies (NB1.1) we used the CLC *de novo* assembler (v4.9), which has also been utilized for the generation of the scarlet macaw and Norway spruce genome assemblies (154, 172). Briefly, the CLC assembler implements the following general procedures: 1) Creation of a table of "words" observed in the sequence data, with retention and utilization of "word" frequency data; 2) Creation of a de Bruijn graph from the "word" table; 3) Utilization of the sequence reads to resolve paths through bubbles caused by SNPs, read errors, and small repeats; 4) Utilization of paired read information (i.e., paired distances and orientation of reads) to resolve more complex bubbles (i.e., larger repeats and/or structural variation); 5) Output of final simple *de novo* contigs (NB1.0) derived from a preponderance of evidence supporting discrete "word" paths, and also supported by the mapping-back process. For the scaffolded *de novo* assembly (NB1.1), the CLC assembler implemented one additional step in which paired reads spanning two contigs were used to estimate the distance between them, determine their relative orientation, and join them where appropriate using "N's"; the

number of which reflect the estimated intercontig distance. Notably, not all *de novo* contigs can be joined to another by read-based scaffolding (i.e., in the absence of map data), and therefore, we use the term scaffolds to collectively refer to the final set of contigs for which read-based scaffolding was attempted. For both assemblies we utilized the same strict assembly parameters in conjunction with all trimmed, unmasked sequence reads (Table 1) as previously described (154), but with the following exceptions: minimum contig length = 300 bp; minimum read length fraction = 0.95; minimum fraction of nucleotide identity (similarity) = 0.95. Paired distances within the Workbench are user-specified, with incorrect specification (i.e., range too narrow or too wide) negatively impacting *de novo* genome assembly. Therefore, using knowledge from library construction and characterization (i.e., agarose gel electrophoresis; Agilent Bioanalyzer) as a guide, we initially assembled the sequence reads multiple times (iteratively), each with incremental increases in the specified paired distances, until the observed paired distances for each library resembled a bell shaped curve centered about a mean that was compatible with library construction and assessment data. For both bobwhite genome assemblies (NB1.0, NB1.1), the user-specified paired distances for all libraries are presented in Table 1. To further suppress genome misassembly, the CLC assembler (i.e., NB1.0, NB1.1) was instructed to break paired reads exhibiting the wrong distance or orientation(s), and only utilize those reads as single reads within the assembly process. This approach is conservative and favors the creation of more contigs with smaller N50 over the creation of larger and fewer contigs that are likely to contain more assembly errors. Assembly statistics for NB1.0 and NB1.1 are provided in Tables S13 and S14 (170).

CHAPTER III

ANNOTATION AND COMPARATIVE ANALYSIS OF THE BOBWHITE GENOME[*]

Introduction

Birds are very unique, they are the only extant phylogenetic class to possess feathers, flight is nearly universal (however not only restricted to birds) with all representatives having or having lost the ability to fly (173), they possess 'small' genome (i.e., one third the size of mammals) (174) and a very distinct karyotype (i.e., $2n \approx 80$, micro and macro chromosomes) (173). Birds are theorized to have originated during the Jurassic from a theropod lineage (i.e., dinosaurs) (175, 176); with the earliest diversification of extant birds (Neornithes) occurring during the Cretaceous (177). Avian lineages have also been noted to exhibit extremely diverse morphologies and rates of diversification across evolutionary time (177). Modern avian species, which consist of approximately 10,500 extant taxa (177), are incredibly diverse in morphology, physiology, and behavior; with many non-domestic species emerging as new animal models with great potential to advance comparative avian genomics, and positively augment conservation efforts for the many threatened bird species (178).

To better understand the genetic complexity of birds, and any links between their genomic variation and phenotypic diversity, a need exists for more avian assemblies (177).

[*]Portions of this chapter were reprinted with permission from "A Draft *De Novo* Genome Assembly for the Northern Bobwhite (*Colinus virginianus*) Reveals Evidence for a Rapid Decline in Effective Population Size Beginning in the Late Pleistocene" by Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, Rollins D, Tizard IR, Brightsmith DJ, Peterson MJ, Taylor JF, Seabury CM, 2014. PLoS ONE 9(3): e90240. doi:10.1371/journal.pone.0090240, Copyright 2014 Halley et al.

Relevant to genome-wide studies of extant wild bobwhite populations, and the utilization of this information to positively augment available management strategies, a need currently exists to generate modern genomic tools and resources such as an annotated bobwhite draft *de novo* genome. Moreover, the impact of utilizing genome-wide polymorphism data for population studies in any species cannot be fully realized without an annotated draft genome assembly. Therefore, we used three *in silico* approaches to facilitate first-generation genome annotation models, and assessed the genomic information content of the draft bobwhite assembly via comparative sequence alignment to the chicken (*G. gallus* 4.0) and zebra finch genomes (*T. guttata* 3.2.4) followed by a genome-wide analysis of divergence (154). The results of this study facilitate genome-wide analyses for the bobwhite, and also enable modern genomics research in other evolutionarily related birds for which research funding is limited.

Results and Discussion

Comparative Genome Alignment

Both bobwhite genome sequence assemblies (i.e., simple contigs, NB1.0; scaffolded assembly, NB1.1) were aligned to the available chicken (*G. gallus* 4.0) and zebra finch (*T. guttata* 3.2.4) reference genomes via blastn (Tables S2 and S3) (170), which allowed for orientation of most *de novo* contigs to their orthologous genomic positions, additional quality control investigations regarding our scaffolding procedure (Table S1) (170), and a genome-wide analysis of divergence with quality control analyses as previously described (154). Examination of the NB1.0 blastn alignments (E-value and

bitscore top hits) across all chicken nuclear chromosomes revealed very stable levels of nucleotide divergence (overall percent identity, Median = 83.20%, Mean = 82.94%), with alignments to GGA24 and GGA16 producing the highest (Median = 85.08%, Mean = 85.05%) and lowest (Median = 76.88%, Mean = 75.48%) percent identities, respectively (Table S2) (170). Evaluation of the NB1.0 blastn alignments (E-value and bitscore top hits) across all zebra finch nuclear chromosomes also revealed stable but greater overall levels of nucleotide divergence (overall percent identity, Median = 77.30%, Mean = 79.04%), with alignments to TGU-LGE22 as well as TGU28 producing the highest (Median ≥ 81.62%, Mean ≥ 81.76%), and TGU16 the lowest (Median = 74.48%, Mean = 75.41%) percent identities, respectively (Table S2) (170). Similar trends in nucleotide divergence were also observed for the NB1.1 blastn alignments to the chicken and zebra finch nuclear chromosomes (Table S3) (170), with greater nucleotide divergence from the zebra finch genome being compatible with larger estimated divergence times (100–106 MYA), as compared to the chicken (56–62 MYA; http://www.timetree.org/) (179, 180).

Annotation of the Bobwhite Genome

Three *in silico* methods were used to annotate the scaffolded bobwhite genome (NB1.1). Initially, we used GlimmerHMM (181, 182) to comparatively predict putative exons within the NB1.1 assembly, with algorithm training conducted using all annotated chicken genes (G gallus 4.0) as recently described (154). The chicken was chosen for training based on the superior level of available annotation and the lowest estimated time since divergence (56–62 MYA), as compared to the zebra finch (100–106 MYA) and the

turkey (56–62 MYA; http://www.timetree.org/) (179, 180). All GlimmerHMM predicted exons were filtered using a high-throughput distributed BLAST engine implementing the blastx algorithm in conjunction with all available bird proteins (NCBI non-redundant avian protein sequences), and the E-value top hits to known avian proteins were retained and summarized (154, 183). Collectively, this comparative *in silico* approach produced statistical evidence for 37,851 annotation models, of which 15,759 represented unique genes and corresponding proteins (Table S6) (170). Similar to the first-generation comparative annotation reported for the scarlet macaw, the number of unique annotation models that are reported here were based on blastx assignments to unique protein hit definitions (i.e. unique accessions), which is known to underestimate the total unique annotation models produced [for review see 154]. As one example, within the NB1.1 assembly, 3,532 genome-wide annotation models were predicted for eight unique protein accessions representing non-LTR retrovirus reverse transcriptases and/or reverse transcriptase-like genes (i.e., pol-like ORFs; RT-like RNA-dependent DNA-polymerases) which have also been predicted in large copy numbers in the chicken nuclear genome (Table S6; GenBank Accessions AAA49022.1, AAA49023.1, AAA49024.1, AAA49025.1, AAA49026.1, AAA49027.1, AAA49028.1; AAA58720.1) (170). Moreover, the prediction of multi-copy genes within all avian genomes routinely utilizes naming schemes which include “like” or “similar to” a specific GenBank accession (154). Our initial comparative annotation procedure culminated with a blastx hit definition representing the highest scoring avian protein curated by NCBI. Therefore, some loci predicted to encode very similar putative proteins, including multi-copy loci such as those

representing gene family members, may be assigned to the same specific protein accession(s) by the blastx algorithm. As occurred for the scarlet macaw genome (154), the absence of bobwhite genome maps and cDNA sequences to guide our initial annotation process also precluded the generation of complete *in silico* models for most bobwhite nuclear genes. Nevertheless, this procedure was successful at identifying bobwhite scaffolds predicted to contain genes encoding moderate to large proteins, which also included some multi-exonic genes distributed across large physical distances (i.e., *TLR2*, *TNRC18*, *NBEA*, respectively; Table S6) (170). Investigation of the blastn comparative alignment data for NB1.1 (Table S3) revealed that all or most of the scaffolds predicted to possess exons encoding these genes (*TLR2*, *TNRC18*, *NBEA*) aligned to their orthologous genomic locations in the chicken (*G. gallus* 4.0) and zebra finch (*T. guttata* 3.2.4) genomes (170). Overall, the results of our comparative annotation for the bobwhite using GlimmerHMM and blastx were similar to those reported for the scarlet macaw (154), but with more annotation models predicted by way of higher genome coverage, and substantially less time since divergence from the chicken.

In a second approach to NB1.1 annotation, we used the Ensembl Galgal4.71 (*G. gallus*) cDNA refseqs (n = 16,396) and ab initio (GENSCAN) sequences (n = 40,571) in an iterative, sequence-based alignment process specifically engineered for transcript mapping and discovery (see Methods; CLC Genomics Large Gap Read Mapper Algorithm, (154)). Of the 56,967 total putative transcripts utilized in this analysis pipeline, 39,603 (70%) were successfully mapped onto the NB1.1 assembly, which included redundant annotation models. Approximately 59% of the mapped transcripts contained gaps which corresponded

to predicted intron-exon boundaries and/or species-specific differences in transcript composition (i.e. regions with no match to NB1.1). Specifically, 12,290 Galgal4.71 cDNA refseq mappings onto NB1.1 were produced, with 10,959 of these possessing unique Ensembl gene names and protein descriptions (Table S7) (170). An additional 27,309 ab initio (GENSCAN) transcripts were also mapped onto NB1.1 (Table S8) (170). An exhaustive summarization of all Galgal4.71 transcript mappings was generated using the sequence alignment map format, and is publicly available (http://vetmed.tamu.edu/faculty/cseabury/genomics). Additionally, the positions of all mapped Galgal4.71 transcripts in NB1.1 and the corresponding gene descriptions (Ensembl, HUGO) are provided in Table S7 (170). Our analysis of these data, including an examination of the scaffolded contig positions (NB1.1) with respect to annotated genes of interest within the chicken genome (*G. gallus* 4.0; Table S7), demonstrates that comparative transcript mapping onto the genomes of more distantly related avian species produces viable annotation models (170). However, this result and corresponding inference is not unique to our study, as other avian genomes (i.e., zebra finch) are often at least partially annotated based on chicken sequences (http://www.ncbi.nlm.nih.gov/genome/367?project _id = 32405).

In a third and final approach to NB1.1 annotation, we utilized the few, low-coverage cDNA sequences that were previously produced for the bobwhite to generate species-specific annotation models. Specifically, we obtained and trimmed 478,142 bobwhite cDNA sequences previously utilized in the construction of a custom bobwhite cDNA microarray (152) (SRA: SRR036708), and subsequently used the quality and

adaptor trimmed reads (n = 325,569; average length = 232 bp) for a strict *de novo* assembly of putative bobwhite transcripts (See Methods). Altogether, 21,367 *de novo* contigs were generated, and of these, 21,011 (98%) were produced from two or more overlapping reads, with most of these contigs (n = 18,135; 85%) possessing ≤ 5X average coverage. Using the same iterative, sequence alignment process (CLC Genomics Large Gap Read Mapper) described for the Galgal4.71 comparative annotation, we successfully mapped 98% of the assembled bobwhite transcripts (n = 21,002) onto NB1.1. Approximately 31% of the mapped transcripts produced gapped alignments that were considered putative intron-exon boundaries. All *de novo* contigs representing bobwhite transcripts were characterized using a high-throughput distributed BLAST engine implementing blastx in conjunction with all available bird proteins (NCBI non-redundant avian protein sequences), and the top ranked hits (i.e., E-value, bitscore) to known avian proteins were retained and summarized (183). Altogether, 8,708 *de novo* contigs (i.e. bobwhite putative transcripts) produced statistical evidence for assignment to at least one known or predicted avian protein (Table S9) (170). Further evaluation of the top hits also revealed some evidence for redundancy across the blastx protein assignments (i.e. same protein; similar alignment length, E-value, and bitscore for two or more avian species). An exhaustive summary of all bobwhite transcript mappings to NB1.1 was also generated using the sequence alignment map format, and is available online (http://vetmed.tamu.edu/faculty/cseabury/genomics). Likewise, the positions of all bobwhite transcripts in NB1.1 are provided in Table S10 (170).

A comparison of all three annotation methods revealed evidence for both novel and redundant annotation models. For example, 8,463 assembled (*de novo*) bobwhite transcripts could be mapped directly onto the Ensembl Galgal4.71 transcripts by sequence similarity and alignment, and of these, 5,537 were redundant with 3,728 unique annotations produced by mapping the Ensembl Galgal4.71 transcripts directly onto NB1.1. Importantly, the overall utility and impact of the previously generated bobwhite cDNA sequences (152) could not be fully realized in the absence of a draft *de novo* genome assembly. Similar to the scarlet macaw genome project (154), both of our bobwhite assemblies (NB1.0, NB1.1) were successful at reconstructing a complete mitochondrial genome at an average coverage of 159X, which resulted in the annotation of 13 mitochondrial protein coding genes (*ND1*, *ND2*, *COX1*, *COX2*, *ATP8*, *ATP6*, *COX3*, *ND3*, *ND4L*, *ND4*, *ND5*, *ND6*, *CYTB*), two ribosomal RNA genes (*12S*, *16S*), 21 tRNA genes, and a predicted D-loop (Table S6) (170). Despite the effectiveness of our mitochondrial and nuclear gene predictions, it should also be noted that even three annotation approaches applied to NB1.1 were not sufficient to exhaustively predict every expected bobwhite nuclear gene. For example, studies of the avian major histocompatibility complex (MHC) have established expectations for gene content among several different bird species, with our approaches providing evidence for many (i.e., *HLA-A*, *TAP1*, *TAP2*, *C4*, *HLA-DMA*, *HLA-B2*, *TRIM7*, *TRIM27*, *TRIM39*, *GNB2L1*, *CSNK2B*, *BRD2*, *FLOT1*, *CIITA*, *TNXB*, *CLEC2D*) but not all previously described avian MHC genes (Table S6) (168-170, 184), (185-189). While the limitations of our three annotation methods were not surprising, the results were sufficient to facilitate informed genome-wide analyses for the bobwhite.

Moreover, even well-established avian genomes, such as the chicken and zebra finch genomes, have yet to be exhaustively annotated. Nevertheless, the results of our annotation analyses provide a foundation for implementing interdisciplinary research initiatives ranging from ecotoxicology to molecular ecology and population genomics in the bobwhite.

Whole-Genome Analysis of Divergence and Development of Candidate Genes

One of the most interesting scientific questions to be directed toward the interpretation of new genome sequences is: "What makes each species unique?" We used the percentile and composite variable approach as well as the validation and quality control procedures previously described (154) to identify *de novo* contigs (NB1.0) displaying evidence of extreme nucleotide conservation and divergence (i.e. outliers) relative to the chicken (*G. gallus* 4.0) and zebra finch (*T. guttata* 3.2.4) genomes (Fig. 2; See Methods). The *de novo* contigs (NB1.0) are useful for this purpose because they provide a shotgun-like fragmentation of the bobwhite genome that is nearly devoid of N's (i.e. intra-contig gaps), which facilitates fine-scale comparative nucleotide alignments that often span large portions, the majority, or even the entire length of the contig sequences. A genome-wide nucleotide sequence comparison of the bobwhite and chicken genomes revealed outlier contigs harboring coding and noncoding loci that were characterized either on the basis of known function and/or the results of human genome wide-association studies (GWAS) (Fig. 2; Table 3; Table S11) [see Appendix A]. Two general trait classes (cardiovascular, pulmonary) were routinely associated with loci predicted within or immediately flanking

the aligned positions of bobwhite contigs (NB1.0) classified as outliers for extreme conservation with the chicken genome (Table 3; Table S11) [see Appendix A]. This result is compatible with the supposition that loci modulating cardiovascular and pulmonary traits are often highly conserved across divergent avian lineages (154). One plausible explanation for this is that birds are unique within the superclass Tetrapoda because they are biologically equipped for both bipedalism and powered flight (190), which may place larger and different demands on the cardiovascular and pulmonary systems than for organisms where mobility is limited to a single terrestrial method (i.e., bipedalism, quadrupedalism). In addition to cardiovascular and pulmonary traits, one bobwhite outlier contig (NB1.0) for extreme conservation with the chicken genome also included a gene (*LDB2*) that is known to be strongly associated with body weight and average daily gain in juvenile chickens (191). This result is compatible with the fact that both the chicken and bobwhite are gallinaceous birds which produce precocial young, and therefore, are likely to share some genetic mechanisms governing early onset juvenile growth and development. Examination of all bobwhite contigs (NB1.0) classified as outliers for divergence with the chicken revealed relatively few predicted genes, with sequences of unknown orthology and noncoding regions being the most common results observed (Table 3; Table S11) [see Appendix A]. This is concordant with the hypothesis that noncoding regions of the genome (i.e., promoters, noncoding DNA possessing functional regulatory elements including repeats) are likely to underlie differences in species-specific genome regulation and traits (192-195). Some of the most interesting bobwhite contigs (NB1.0) displaying evidence for extreme divergence were predicted to contain putative introns for *CSMD2* as well as *TNIK*,

and to flank *LPHN3* (intergenic region; Table 3; Table S11) [see Appendix A]. These three genes have all been associated with human brain-related traits including heritable differences in brain structure (*CSMD2*, voxel measures) (196) measures of activation within the dorsolateral prefrontal cortex (*TNIK*) (197) and working memory in schizophrenia patients receiving the drug Quetiapine (198). Our whole genome-wide analysis of divergence between the bobwhite and the chicken provides further evidence that noncoding regions of the genome are likely to play a tangible role in the developmental manifestation of species-specific traits (192-195), including both neurocognition and behavior (196-198).

Comparison of the bobwhite (NB1.0) and zebra finch genomes (*T. guttata* 3.2.4) also revealed evidence for extreme nucleotide conservation and divergence (Fig. 2; Table 4; Table S11) [see Appendix A]. In comparison to the zebra finch genome, two general trait classes (osteogenic, cardiovascular) were routinely associated with loci predicted within or immediately flanking the aligned positions of bobwhite contigs (NB1.0) classified as outliers for extreme conservation (Table 4; Table S11) [see Appendix A]. Within these contigs, the presence of orthologous gene sequences previously associated with human cardiovascular traits (or their proximal noncoding flanking regions) was relatively unsurprising, as this result also occurred during our analysis of divergence with the chicken genome (Table 3; Table 4; Table S11) [see Appendix A], and in a previous study of the scarlet macaw genome (154). Therefore, it is apparent that some loci associated with cardiovascular and pulmonary traits in humans appear to be extremely conserved across multiple avian species, including some of the same loci identified by

similar analyses involving the scarlet macaw, chicken, and zebra finch genomes (Table S11) (154). Among the bobwhite contigs classified as outliers for extreme conservation with the zebra finch, we also observed orthologous gene sequences (or their proximal noncoding flaking regions) which were previously associated with human bone density, strength, regeneration, and spinal development as well as human height and waist circumference (Table 4; Table S11) [see Appendix A]. Interestingly, the overall size and stature of the bobwhite (i.e. height or length, wingspan) is actually more similar to the zebra finch than to the chicken (10, 199, 200), which is compatible with these results. Additionally, while the temporal order of ossification for avian skeletal elements is known to be conserved across divergent bird species (i.e., duck, quail, zebra finch) (201), some aspects of wild bobwhite medullary bone formation (i.e., annual frequency of occurrence) are arguably far more similar to the zebra finch than to domesticated chickens, which have been bred and utilized for continuous egg production (202-204). Therefore, some similarities in the underlying biology of these two bird species were reconciled with the genomic information content found within several bobwhite outlier contigs displaying evidence for extreme conservation with the zebra finch genome. At the opposite end of the distribution (Fig. 2), and across all diverged outliers with respect to the zebra finch genome, one of the most intriguing results was a bobwhite contig predicted to contain an *LDB2* intron (Table 4; Table S11) [see Appendix A]. Notably, *LDB2* was implicated as an outlier for extreme conservation with the chicken genome (Table 3; Table 4; Table S11) [see Appendix A], and is known to be strongly associated with body weight and average daily gain in precocial juvenile chickens (191). The observation of this same putative gene

(a different NB1.0 contig) with respect to extreme divergence with the zebra finch genome (Table 4; Table S11) [see Appendix A] may potentially reflect the different developmental strategies associated with the bobwhite and the zebra finch (i.e., precocial versus altricial) (58, 205, 206). Two additional contigs classified as outliers for divergence were also predicted to be proximal to genes implicated by human GWAS studies for age at menarche (*NR4A2*) (207) and reasoning in schizophrenia patients receiving the drug Quetiapine (*ZNF706*; Table 4; Table S11) [see Appendix A] (198). Interestingly, both wild and domesticated zebra finches reach sexual maturity earlier than do bobwhites, with hypersexuality in the zebra finch considered to be an adaptation to arid environments (208, 209). However, any potential relationships between *ZNF706* and specific underlying biological differences between the bobwhite and zebra finch were not apparent, especially since no studies have comparatively evaluated a battery of cognitive traits in these two species using standardized methods.

Table 3. Biologically Relevant NB1.0 Simple *De Novo* Outliers from a Genome-wide Analysis of Divergence with the Chicken Genome (*G. gallus* 4.0).

Predicted Outlier Contig Genes[a,b,c]	**Known Function or GWAS Trait Classification**	**References**
BCL11B[a]	Aortic Stiffness	(210)
ALPK3[a]	Cardiac Heath and Development	(211)
SETBP1[a], *FAF1*[a]	Heart Ventricular Conduction	(212)
MEF2A[a], *LPL*[a],	Cardiomyopathy	(213, 214)
KCNJ2[a]	Heart Q-wave T-wave Interval Length	(215)
LDB2[a], *PTPRF*[a],	Coronary Artery Disease	(216-218)
ATP10B[a]		
ZNF652[a], *FIGN*[a],	Blood Pressure	(219, 220)
CHIC2[a]		
CFDP1[a], *KCNJ2*[a]	Pulmonary Function and Health	(221, 222)
GRM3[a], *RELN*[a], *RORB*[a],	Cognitive Abilities	(223-225)
CSMD2[b]	Brain Structure	(196)
TNIK[b]	Brain Imaging	(197)
LPHN3[b]	Working Memory	(198)

[a] Outlier for extreme nucleotide-based conservation.
[b] Outlier for extreme nucleotide-based divergence.
[c] See Supplemental Table S11 for an exhaustive list of outlier contigs with annotation.

Table 4. Biologically Relevant NB1.0 Simple *De Novo* Outliers from a Genome-wide Analysis of Divergence with the Zebra Finch Genome (*T. guttata* 3.2.4).

Predicted Outlier Contig Genes[a,b,c]	**Known Function or GWAS Trait Classification**	**References**
CDH13[a], *CXADR*[a],	Blood Pressure	(226, 227)
VTI1A[a], *KLF12*[a]	Heart Ventricular Conduction	(212)
BCL11B[a]	Aortic Stiffness	(210)
GJA1[a]	Resting Heart Rate	(228)
JAG1[a]	Bone Density	(229)
VPS13B[a]	Bone Strength	(230)
SALL1[a]	Bone Mineral Density	(231)
STAU2[a]	Spinal Development	(232)
SATB2[a]	Osteogenic Differentiation And Regeneration	(233)
ZFHX4[a], *BNC2*[a],	Height	(234, 235)
STX16[a], *APCDD1L*[a]	Waist Circumference	(236)
GRIA1[a]	Anthropometric Traits	(237)
LDB2[b]	Body Weight	(191)
LDB2[b]	Average Daily Gain	(191)
NR4A2[b]	Age of onset of Menarche	(207)
ZNF706[b]	Reasoning	(198)

[a] Outlier for extreme nucleotide-based conservation.
[b] Outlier for extreme nucleotide-based divergence.
[c] See Supplemental Table S11 for an exhaustive list of outlier contigs with annotation.

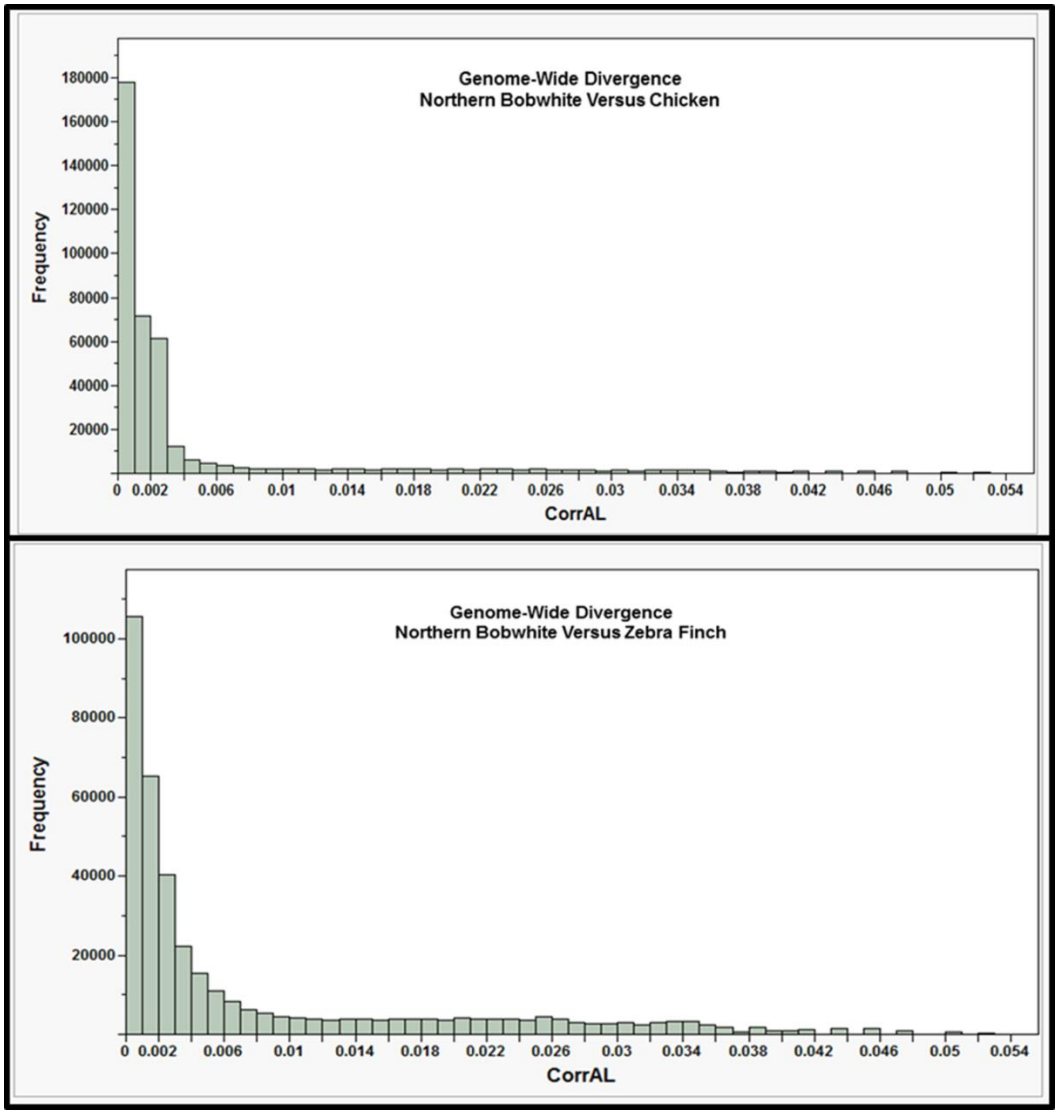


Figure 2. Whole Genome Analysis of Divergence. (Top) Genome-wide nucleotide-based divergence (CorrectedForAL) between the bobwhite (*Colinus virginianus*; NB1.0; simple *de novo* assembly) and the chicken genome (*Gallus gallus* 4.0). (Bottom) Genome-wide nucleotide-based divergence (CorrectedForAL) between the bobwhite (*Colinus virginianus*; NB1.0; simple *de novo* assembly) and the zebra finch genomes (*Taeniopygia guttata* 1.1, 3.2.4). Each histogram represents the full distribution of the composite variable defined as: CorrectedForAL = $\frac{(\frac{PercentID}{100})}{AlignmentLength}$ (154). The left edges of the distributions represent extreme conservation, whereas the right edges indicate extreme putative divergence. The observed ranges of the composite variable were 2.19545E-05 – 0.052631579 (chicken), and 4.28493E-05 – 0.052631579 (zebra finch). Distributional outliers were predicted using a percentile-based approach (99.98th and 0.02th) to construct interval bounds capturing > 99 % of the total data points in each distribution.

Quality Control Investigation for Analyses of Divergence

All NB1.0 contigs classified as putative outliers for divergence (Fig. 2; right tail) shared one unifying feature: A 19–20 bp alignment with 100% identity to a reference genome (i.e., chicken or zebra finch) regardless of contig size (Range = 300 bp to 1,471 bp; Median = 385 bp; Mean = 438 bp). These short alignments had variable sequences, with the common feature being the short length (19–20 bp), and produced values for the composite variable (CorrectedForAL $=\frac{\left(\frac{PercentID}{100}\right)}{Alignment\ Length}$) that ranged from 0.050 to 0.053 (i.e., $\frac{\frac{100}{100}}{10\text{bp}}, or\ \frac{\frac{100}{100}}{19\text{bp}}$). This was expected based upon previous observations (154), and at least three plausible explanations for this result include: 1) The orthologous sequences are simply missing from the chicken and/or zebra finch genome assemblies; 2) The NB1.0 contigs are misassembled; or 3) The NB1.0 contigs represent true outliers for nucleotide divergence and include species-specific insertion-deletion mutations. Some sequences are invariably missing from every draft genome assembly (i.e., unassembled). Therefore, we searched five databases curated by NCBI (i.e., refseq_genomic, refseq_rna, nr/nt, traces-WGS, traces-other DNA) for nucleotide alignments that would facilitate NB1.0 contig characterization and/or help refute the diverged outlier status of these contigs, and in all cases found little or no evidence for a conclusively better blastn alignment to the chicken or zebra finch genomes (See Methods). However, some of these contigs actually produce better blastn alignments (i.e., E-value, bitscore) to other vertebrate species, including other avian species, which is not compatible with outlier status (diverged) resulting solely from contig misassembly (Table S2; Table S11).

Regarding our whole-genome analyses of divergence, all NB1.0 contigs classified as outliers for extreme conservation (Fig. 2; extreme left edge) were moderately large (Range = 9,647 bp to 89,591 bp; Median = 22,792 bp; Mean = 25,196 bp) in comparison to outliers for divergence (Range = 300 bp to 1471 bp; Median = 385 bp; Mean = 438 bp). Again, this trend was expected and has been previously described (154). Therefore, we conducted several quality control (QC) analyses that were designed to assess whether factors other than nucleotide sequence divergence were responsible for our results. First, we used summary data from the two comparative genome alignments performed using blastn to estimate pairwise correlations among the following: NB1.0 contig size (bp), contig percent GC, contig percent identity, and contig alignment length (bp). Moderate correlations between NB1.0 contig alignment length and contig size were observed with respect to the chicken ($r = 0.649$, Nonparametric $\tau = 0.656$) and zebra finch genome alignments ($r = 0.490$, Nonparametric $\tau = 0.492$), whereas weak correlations were observed between percent identity and alignment length (chicken: $r = 0.127$, Nonparametric $\tau = 0.071$; zebra finch: $r = -0.371$, Nonparametric $\tau = -0.469$). Weak correlations were also observed for all other investigated parameters. This result is important because the two parameters that drive our analysis of divergence are the percent identity and the alignment length, which were jointly used to construct a composite variable (CorrectedForAL) representing percent identity normalized for alignment length across all NB1.0 contigs which produced blastn alignments to the chicken and zebra finch genomes. In a second QC analysis, we applied the same percentile based approach (Percentiles = 99.98th and 0.02th) used in our whole-genome analyses of divergence to examine the full, ordered distribution

of NB1.0 contig sizes, and determined that only 2 contigs (chicken analysis; contigs 4309, 7216) were in common with the 244 implicated as outliers for conservation or divergence (Table S11). This result argues against contig size being deterministic for outlier status. Finally, for larger contigs, such as those classified as outliers for conservation, the blastn procedure often produces multiple meaningful alignments, which are appended below the most "significant" hit (i.e., E-value and bitscore top ranked hit). These appended alignments include both noncontiguous (i.e., gaps due to insertion-deletion mutations) and less "significant" comparative alignments (i.e., increasing nucleotide sequence divergence). To assess the reliability of utilizing only the top ranked hit (i.e., E-value and bitscore) as a proxy for larger contigs which may produce multiple, syntenic, noncontiguous hits spanning either the majority or even the entire contig length, we used the additional (i.e., appended) non-overlapping alignment data (percent identity, alignment length) for the conserved outlier contigs to recalculate our composite variable (Table S12) (170). Across all 145 unique contigs categorized as conserved outliers, the new (recalculated) composite variable only further confirmed the original outlier status (i.e., extreme conservation), which is in agreement with the results of a similar study involving the scarlet macaw genome (Table S12) (154, 170). Moreover, the NB1.0 contigs classified as outliers for extreme conservation are actually highly conserved genomic regions for which extended nucleotide conservation persists for the two compared species, which cannot occur in the presence of species-specific genomic rearrangements, copy number variants whereby one or more amplification-deletion boundaries are traversed, or in the presence of frequent and complex repetitive elements. Nevertheless, only NB1.0 contigs

which produced blastn results (> 99%) could be included in our analyses of divergence and quality control analyses, as they provided the data required to construct the composite variable. All NB1.0 contigs for which no alignments were achieved with respect to the chicken or zebra finch genomes are provided in Table S2 (170).

Methods

Estimating Concordance between Genome Assemblies

Treating all NB1.0 contig sequences as individual sequence reads, we used the CLC Large Gap Read Mapper algorithm to iteratively search the scaffolded genome assembly (NB1.1) for the best matches (v2.0 beta 10) as previously described (154). A single, initial round of iterative searching resulted in 91% of the NB1.0 contigs mapping onto the NB1.1 assembly, with 99% of these mappings containing no gaps. Thereafter, a SAM output was created, which was then used to parse out the coordinates of all mapped NB1.0 contigs for the purpose of creating a reference table summarizing the concordance between the two assemblies (Table S1) (170). All parsing and joining was performed using Microsoft SQL Server 2008 R2.

Comparative Genome Alignment

The NB1.0 and NB1.1 genome assemblies were aligned to the chicken (*G. gallus* 4.0) and zebra finch (*T. guttata* 1.1, 3.2.4) reference genome assemblies (including ChrUN, unplaced) using the blastn algorithm (version 2.2.26+). To minimize disk space and enable continuous data processing we used an E-value step-down procedure as recently described

(154). After each step, we exported the results and parsed out the top hit (E-value, bitscore) for each bobwhite contig (NB1.0, NB1.1). E-value ties were broken by bitscore. All parsing was performed using Microsoft SQL Server 2008 R2.

*"*In silico*" Annotation of the Bobwhite Genome*

Initially, we used GlimmerHMM (169, 181, 182) to predict exons and putative gene models within NB1.1. GlimmerHMM was trained using all annotated chicken genes (G gallus 4.0) as recently described (154), which is similar to an approach used for annotation of the turkey genome (169). Thereafter, we characterized, assessed support, and filtered GlimmerHMM predictions via blastx (183) in conjunction with all available bird proteins (NCBI non-redundant avian protein sequences), with the top hits (E-value, bitscore; minimum E-value = 1E-04) to known avian proteins retained and summarized as previously described (154).

In a second approach to annotation, we used the Ensembl Galgal4.71 (*G. gallus*) cDNA refseqs (n = 16,396) and ab initio (GENSCAN) sequences (n = 40,571) in an iterative, sequence-based alignment process for comparative transcript mapping and discovery. Galgal4.71 transcript length ranged from 108 bp to 93,941 bp. Briefly, we used the CLC large gap read mapper (v2.0 beta 10) to iteratively search the NB1.1 assembly for the best Galgal4.71 nucleotide matches. The CLC large gap read mapper was utilized as previously described (154), but with the following exceptions: maximum distance from seed = 100,000; minimum fraction of identity (similarity) = 0.80; minimum read length fraction = 0.001. Our settings for minimum read length fraction were necessary to facilitate

mapping for large Galgal4.71 transcripts. However, this setting did not impede or nullify the stringency of mapping smaller transcripts, as the best matches (i.e. longest length fraction and highest similarity) were sought and reported. A SAM file representing all Galgal4.71 mappings was created using the CLC Genomics Workbench. Gene names (HUGO), descriptions, and protein information for the Ensembl Galgal4.71 cDNA refseqs were obtained from BioMart-Ensembl (http://useast.ensembl.org/biomart/martview/) and NCBI (http://www.ncbi.nlm.nih.gov/sites/batchentrez).

In a third approach to annotation, we obtained 478,142 bobwhite cDNA sequences (Roche 454) previously used to construct a microarray (152) (SRA: SRR036708) and trimmed them for quality and adaptors. Thereafter, the remaining sequences (n = 325,569; average length = 232 bp) were assembled using the CLC *de novo* assembler (v6.0.4) and the same strict assembly parameters utilized for NB1.0 and NB1.1. *De novo* contigs (50 bp to 6466 bp) generated from bobwhite cDNA sequences were mapped onto NB1.1 using the CLC large gap read mapper as described above for the Galgal4.71 transcripts, but with the following modifications: minimum fraction of identity (similarity) = 0.90; minimum read length fraction = 0.01. All *de novo* contigs generated from bobwhite cDNA sequences were characterized using blastx (183) in conjunction with all available bird proteins (NCBI non-redundant avian protein sequences) as previously described (154). A SAM file representing all bobwhite cDNA *de novo* contig mappings was created using the CLC Genomics Workbench.

The bobwhite contig containing the mitochondrial genome (NB1.0, NB1.1) was manually annotated using the chicken as a guide (GenBank Accession HQ857212), and

several available BLAST tools (blastn, bl2seq, blastp; http://blast.ncbi.nlm.nih.gov/). Thereafter, we used tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) to predict tRNA genes, with one tRNA manually predicted by comparative sequence analysis.

Whole-Genome Analyses of Divergence and Development of Candidate Genes

For all NB1.0 contigs that produced blastn hits to the chicken (*G. gallus* 4.0) or zebra finch genomes (*T. guttata* 3.2.4), we normalized the observed percent identity for differences in alignment length across both comparative genome alignments using the following formula: CorrectedForAL = $\frac{\frac{PercentID}{100}}{Aligment\ Length}$ (154). This method is mathematically similar and related to the p-distance (238), and allows for genome-wide nucleotide by nucleotide comparison of both coding and noncoding DNA, with a previous investigation supporting the use of alignment based sequence comparison and distance estimation for conserved genomes (239). Thereafter, we visualized the full distribution of this composite variable by producing histograms within JMP Pro 10.0.1 (SAS Institute Inc., Cary, NC). The full distribution of observed "CorrectedForAL values" produced from each comparative genome alignment is highly skewed and resistant to standard transformation methods (154). Therefore, we used a percentile approach to identify outlier contigs based on establishing interval bounds within the ordered distributions (at the 99.98^{th} and 0.02^{th} percentiles). All analytical procedures including outlier definition, detection by percentile-cutoff locations, and quality control analyses followed methods previously described (154). All NB1.0 contigs implicated as outliers for divergence were scrutinized by searching five databases curated by NCBI (i.e., refseq_genomic, refseq_rna,

nr/nt, traces-WGS, traces-other DNA) for blastn alignments that would further confirm or refute their outlier status. Trace alignments (i.e., WGS; other) with bitscores ≥ 15% larger than the original bitscore were considered false positives for extreme divergence, and were removed from the final list of putative outliers. NB1.0 contigs classified as outliers for extreme conservation were annotated based on the individual reference genome from which they were identified (i.e., *G. gallus* 4.0; *T. guttata* 3.2.4; See Table S11) (170). Established knowledge of gene function (i.e., among outliers) in combination with the human GWAS literature were used to identify potential candidate genes for biological traits among the avian species compared.

CHAPTER IV

REPEAT CONTENT, GENOME-WIDE VARIANT DETECTION AND BOBWHITE POPULATION HISTORY INFERRED FROM WHOLE-GENOME SEQUENCE DATA*

Introduction

Modern DNA sequence data can actually reveal a lot about ancient population history, since the genome of an extant individual is a mosaic of the genomes of its ancestors (240). Estimates of the effective population size seek to consider only those individuals in a population that pass on genes to the next generation, and is a central metric in the field of population genetics (241, 242). The pairwise sequentially Markovian coalescent (PSMC) is a method that was originally developed to quantify historical variation in effective population sizes in humans (243), and has demonstrated its utility in various other species (242, 244-246). Avian model systems have had a rich history of research focusing on patterns of diversity and historical biogeography (241, 247, 248); with one recent study utilizing 38 individual bird genomes to quantify how population sizes of avian species have changed globally over the past few million years (242). Analyses performed using PSMC, while not very reliable in relation to recent population history of a species (i.e., less than a few thousand years) (243), do have important implications for the conservation of extant birds (241).

* Portions of this chapter were reprinted with permission from "A Draft *De Novo* Genome Assembly for the Northern Bobwhite (*Colinus virginianus*) Reveals Evidence for a Rapid Decline in Effective Population Size Beginning in the Late Pleistocene" by Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, Rollins D, Tizard IR, Brightsmith DJ, Peterson MJ, Taylor JF, Seabury CM, 2014. PLoS ONE 9(3): e90240. doi:10.1371/journal.pone.0090240, Copyright 2014 Halley et al.

Birds exhibit a broad spectrum of developmental patterns (i.e., rates of maturation, behavior, physiology, anatomy) (58, 60, 249), with a direct correlation between functional maturity of the chick, and the level of care it receives from its parents (58). These differences in development have lead scientists to characterize birds based on an altricial-precocial development spectrum (58). The bobwhite is an example of a precocial bird (i.e., chicks can fend for themselves and are self-thermoregulatory) (58, 60) while the scarlet macaw is an example of an altricial bird (i.e., depend on parents for food and thermoregulation) (250). Additionally, because of the developmental differences between altricial and precocial birds, precocial chicks have a higher likelihood of survival post-death of parental caregivers than do altricial chicks (58).

In addition to being precocial, the bobwhite exhibits a variety of biological traits associated with *r*-selection (i.e., produce many offspring that have a low probability of survival, early maturity), while the scarlet macaw displays clear evidence for being a *K*-selected (i.e., produce few offspring who have a higher life expectancy, late maturity) avian species (57, 251, 252). For this reason, we will use these two species as an experimental *in silico* model to test the hypothesis that the historic effective population size (N_e) for an *r*-selected avian species will exceed that of a *K*-selected avian species, across all relevant time points, and then compare the magnitude by which they differ. To reconstruct and compare the demographic histories of the bobwhite and the scarlet macaw the PSMC model will be used (243).

Results and Discussion

Predicted Repeat Content, and Genome-Wide Variant Detection

The minimum estimated repetitive DNA content (excluding N's) for the scaffolded bobwhite genome was approximately 8.08%, as predicted by RepeatMasker (RM; Table 5; Table S4) (170). This estimate was greater than those reported for the Puerto Rican parrot, saker and peregrine falcon, scarlet macaw, turkey, and zebra finch genomes using RM (154, 161, 163, 169, 184), but less than that reported for the chicken genome (168). However, read-based scaffolding involving the insertion of "N's" into gaps is known to result in the underestimation of genome-wide repetitive content (154). Nevertheless, a common feature of the bobwhite, scarlet macaw, chicken, turkey, and zebra finch genomes is the high proportion of LINE-CR1 interspersed repeats (154, 168, 169, 184) that are conserved across these divergent avian lineages. In fact, the majority of the predicted repeat content in the bobwhite genome consisted of interspersed repeats, of which most belong to four groups of transposable elements including SINEs, L2/CR1/Rex non-LTR retrotransposons, retroviral LTR retrotransposons, and at least three DNA transposons (hobo Activator, Tc1-IS630-Pogo, PiggyBac). Similar to the chicken, the bobwhite genome was predicted to contain about one third as many retrovirus-derived LTR elements as the zebra finch (184), but more SINEs than the chicken (168, 184). To further evaluate the repetitive content within the bobwhite genome, we utilized PHOBOS (v3.3.12) (87) to predict and characterize genome-wide tandem repeats (microsatellite loci) for the purpose of identifying loci that could be utilized for population genetic studies. Collectively, we

identified 3,584,054 tandem repeats (Table S5) consisting of 2 to 10 bp sequence motifs that were repeated at least twice, which is greater than 50% more tandem repeats than was recently predicted for the scarlet macaw (154, 170). Bobwhite tandem repeats were characterized as follows: 644,064 di-, 997,112 tri-, 577,913 tetra-, 518,315 penta-, 552,957 hexa-, 143,590 hepta-, 93,583 octa-, 35,260 nona-, and 21,260 decanucleotide microsatellites (Table S5) (170). Importantly, microsatellite genotyping as a means to assess parentage, gene flow, population structure, and covey composition within and between bobwhite populations has historically been limited to very few genetic markers (102, 253, 254), and therefore, the resources described herein will directly enable genome-wide population genetic studies for the bobwhite.

To provide the first characterization of genome-wide sequence variation for a wild bobwhite, we investigated the frequency and distribution of putative single nucleotide polymorphisms (SNPs) and small insertion-deletion mutations resulting from biparental inheritance of alternative alleles (heterozygosity) within the repeat-masked scaffolded *de novo* assembly (NB1.1). Collectively, 3,503,457 SNPs and 268,981 small indels (Coverage $\geq 10\times$ and $\leq 572\times$) were predicted (Fig. 3), which corresponds to an average genome-wide density (i.e., intra-individual variation) of approximately 3.22 heterozygous polymorphisms per Kbp for the autosomes. Considering only high quality putative SNPs, the bobwhite heterozygous SNP rate was approximately 2.99 SNPs per Kbp. This estimate is four times greater than that reported for the peregrine falcon, more than three times greater than for the scarlet macaw and saker falcon, approximately twice that of the zebra finch and turkey, and is second only to the chicken and the flycatcher, which are most

similar to the bobwhite in terms of putative heterozygous SNPs per Kbp (154, 162, 163, 169, 184, 255). Despite evidence for recent population declines across the majority of the bobwhite's historic U.S. range (73, 74, 77-80), our wild Texas bobwhite possesses extraordinary levels of genome-wide variation as compared to most other avian species for which draft *de novo* genome assemblies are currently available.

Table 5. Major Classes of Repetitive Content Predicted by RepeatMasker within the NB1.1 Scaffolded *De Novo* Assembly.

Repeat Type Predicted	**Total Elements[a]**	**Total bp (% of Genome)[a]**
SINEs	4,425	545,252 (0.047%)
LINEs (L2/CR1/Rex)	172,398	44,762,255 (3.818%)
LTR Retroviral	31,766	8,987,247 (0.767%)
DNA Transposons	22,793	6,863,495 (0.585%)
Unclassified Interspersed Repeats	2,096	337,844 (0.0288%)
Small RNA	757	70,666 (0.006%)
Satellites	3,624	580,253 (0.050%)
Low Complexity & Simple Repeats	403,599	32,608,785 (2.781%)
Totals	**641,458**	**94,755,797 (8.08%)**

[a]Scaffolded *de novo* assembly NB1.1 (1.17 Gb including gaps with N's

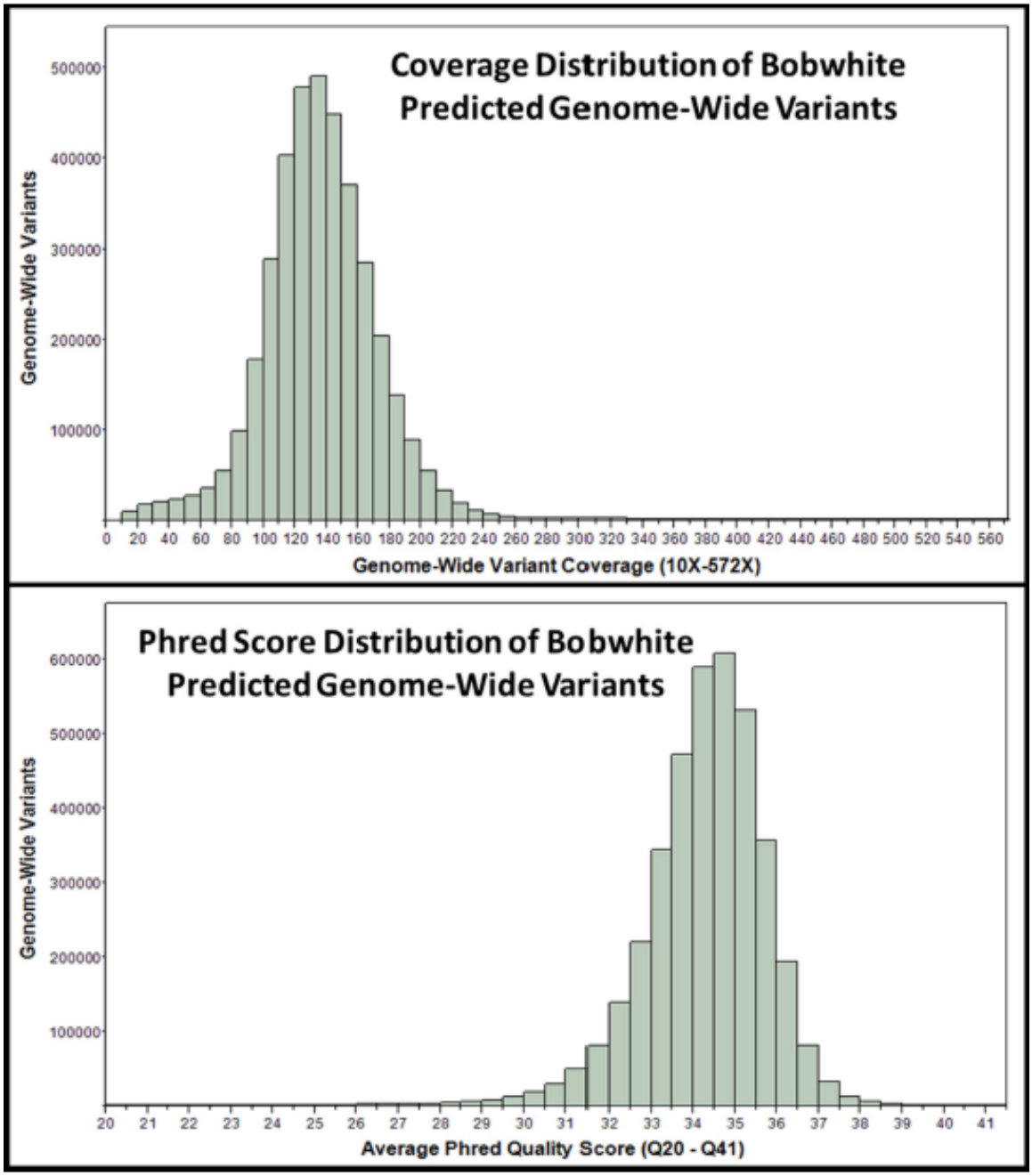


Figure 3. Autosomal Coverage and Quality Score Distributions for Variants Predicted in the Scaffolded Bobwhite (*Colinus virginianus*) Genome (NB1.1). Total genome-wide variants predicted within NB1.1 appears on the y-axis, with coverage and quality scores presented on the x-axis, respectively. Total variants include putative single nucleotide polymorphisms and small insertion deletion mutations (≤ 5 bp) that were predicted within the repeat masked NB1.1 assembly.

Bobwhite Population History as Inferred From Whole-Genome Sequence Data

Using high-quality autosomal SNP density data, we implemented a pairwise sequentially Markovian coalescent (PSMC) model (243) to reconstruct the demographic history of our wild bobwhite (Pattie Marie), and for comparison, we also produced a PSMC analysis for a wild female scarlet macaw (Neblina; Fig. 4) (154). For both species, we inferred their demographic history using the per-site pairwise sequence divergence to represent time, and the scaled mutation rate to represent population size (243). Importantly, many biological characteristics associated with the bobwhite are largely typical of an *r*-selected avian species, whereas the scarlet macaw clearly exhibits characteristics of *K*-selection (2, 57, 251, 252). However, despite the fundamental biological differences in how these two avian species achieve reproductive success within their respective habitats, both species experienced pronounced bottlenecks which were predicted to begin approximately 20–58 thousand years ago (kya), with the range in timing of this interval being a product of modeling a range of underlying mutation rates (Fig. 4; See Methods). The temporal synchronicity of these bottlenecks for the bobwhite and the scarlet macaw became more coincident as the assumed mutation rate approached the human mutation rate (PSMC default $\mu = 2.5\times10^{-8}$). Beginning approximately 20 kya, the bobwhite (generation time=1.22 yrs; Fig. 4) and the scarlet macaw (generation time = 12.7 yrs; Fig. 4; See Methods) demonstrate synchronous declines in their estimated effective population sizes (N_e), with this trend persisting up until about 9–10 kya, which is coincident with the timing of modern human colonization of the New World (15,500–40,000 years ago) (256-259), the collapse of the megafauna (260-262), and the last glacial maximum (LGM) (263, 264).

The geographic expansion of modern man has previously been proposed (i.e., subsistence hunting; overkill) as one highly efficient mechanism for the late Pleistocene collapse of the megafauna in the Americas, and to a lesser degree, in Eurasia (260, 262). Both the bobwhite and the scarlet macaw were hunted by indigenous peoples of the Americas (5, 265-267). However, the peregrine falcon also experienced a bottleneck at about the same time as the bobwhite and the scarlet macaw, possibly due to climate-driven habitat diminution (163), which may also explain some or even most aspect(s) of the predicted declines that we detected. Moreover, the peregrine falcon previously used for PSMC modeling was not sampled from the New World (163), which further confirms the possibility for the LGM (263, 264) being explanatory for temporally relevant global declines of many animal populations, with recent evidence of swine population declines (i.e., European and Asian wild boar; *Sus scrofa*) (244) during the same time intervals as the bobwhite and scarlet macaw declines (Fig. 4).

Relevant to modern conservation biology and conservation genetics, it is clear that the estimated N_e of the bobwhite remained large even after a historic bottleneck (i.e., up to about 9–10 kya), with a historic peak N_e which was more than 6.6 times larger than the scarlet macaw (Fig. 4). This result was relatively unsurprising given the high autosomal SNP rate predicted for the bobwhite in this study (2.99 SNP per Kbp). When avian mutation rates (i.e., bobwhite, scarlet macaw) were modeled according to the human mutation rate (PSMC default $\mu = 2.5\times10^{-8}$), as was also assumed for the wild boar (244), peak N_e for the bobwhite was estimated at approximately 95,000 about 20 kya, with a subsequent decline to approximately 72,000 by 9–10 kya (Fig. 4). The most recent

bobwhite peak which arises near 10^{-4} on the "Time" x-axis (scaled in units of 2μT) appears to be an artifact due to PSMC being unable to model a continued decline in N_e until the present, with a similar statistical signature and corresponding overestimation of N_e detected prior to a population decrease that was predicted in the Denisovan genome analysis (268). Estimates of modern N_e in the bobwhite will require multiple sequenced individuals (240) to adequately estimate the severity of the predicted decline. Relevant to modern bobwhite declines observed across the majority of their U.S. range (73, 74, 77-80), our demographic analysis indicates that the *r*-selection strategy employed by the bobwhite can be very effective with respect to rapid increases in N_e (i.e., see the increase at 4×10^{-3} 2μT in Fig. 4). Therefore, it is apparent that these recent bobwhite declines may potentially be reversed at least to some degree (i.e., boom-bust pattern) in regions with suitable habitats, ample annual rainfall, and low harvest intensity. In striking contrast to the bobwhite, peak N_e for the scarlet macaw (assuming $\mu = 2.5\times10^{-8}$) was never as large, and was estimated at approximately 15,500 about 25 kya, with a subsequent collapse to approximately 3,000 by 2.5 kya (Fig. 4); despite the fact that Neblina is from Brazil (i.e., wild caught) and was part of the population found in the Amazon Basin and adjacent lowlands, with an estimated population habitat range that exceeds 5 million km^2. Our analysis of these data strongly underscores the importance of conservation biology and conservation genetics in the scarlet macaw and other related pscittacines that rely heavily on *K*-selection (57, 251, 252). Notably, the disparities in peak N_e as well as the more recent estimates (10 kya) for the bobwhite and the scarlet macaw are likely to reflect long-term, opposing differences in the *r*-/*K*- selection continuum (57, 251, 252), and suggest that

species which rely heavily on facets of *K*-selection for success, like the scarlet macaw, could be at higher risk of experiencing more rapid and dramatic declines in N_e that are likely to prolong recovery. In fact, even under the perception of relatively ideal biological conditions in the field, N_e for large *K*-selected avian species like the scarlet macaw may be much lower than presumed based on the amount of available habitat, and the estimated total population size. Our findings highlight the need to conserve large populations of scarlet macaws and similar species in order to maintain genomic diversity and corresponding N_e to avoid unmasking deleterious alleles by way of increasing homozygosity, as observed for the highly endangered Spix's Macaws (269, 270). However, caution is necessary when interpreting the results of PSMC, as population size reductions and population fragmentation may not always be easily differentiated (243).

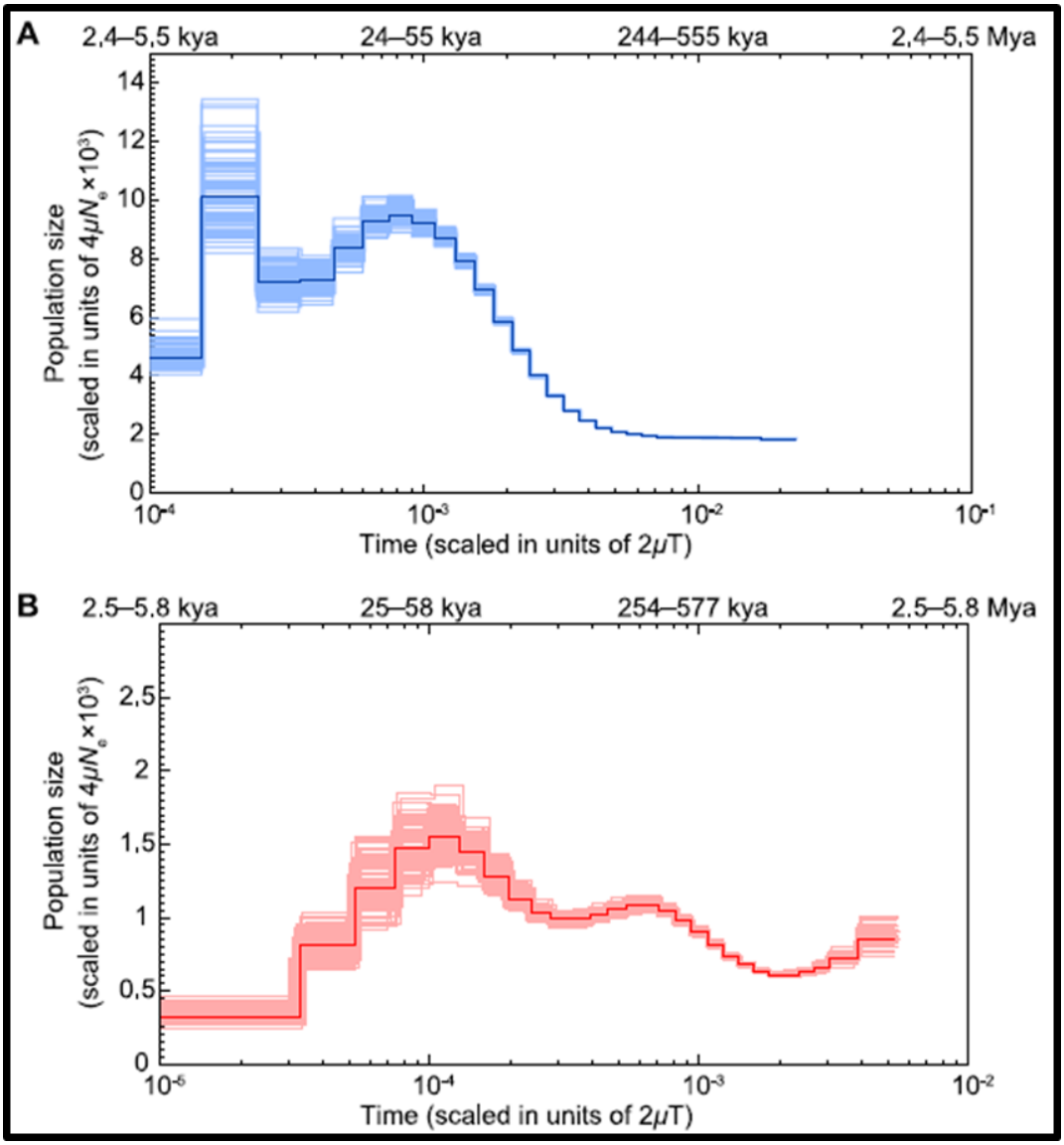


Figure 4. Comparative Demographic History Analysis and PSMC Effective Population Size Estimates for Bobwhite (*Colinus virginianus*) (A) and Scarlet Macaw (*Ara macao*) (B). Estimates of effective population size are presented on the y-axis as the scaled mutation rate. The bottom x-axis represents per-site pairwise sequence divergence and the top x-axis represents years before present, both on a log scale. Generation intervals of 1.22 years for the bobwhite (*Colinus virginianus*) and 12.7 years for the scarlet macaw (*Ara macao*) were used (See Methods). In the absence of known per-generation *de novo* mutation rates for the bobwhite and the scarlet macaw, we used the two human mutation rates (μ) of 1.1×10^{-8} and 2.5×10^{-8} per generation (271, 272) (see Methods). Darker lines represent the population size inference, and lighter, thinner lines represent 100 bootstraps to quantify uncertainty of the inference.

Methods

Characterization of Repeat Content and Variant Prediction

To estimate the minimum repetitive content within the bobwhite genome (NB1.1), we processed all of the scaffolds with RepeatMasker (http://www.repeatmasker.org/; RepBase16.0.1). As described for the scarlet macaw genome (154), we conducted a two-stage, composite analysis which consisted of masking the NB1.1 contigs with both the chicken and zebra finch repeat libraries to cumulatively estimate the detectable repetitive content. Additionally, we used PHOBOS (v3.3.12) (273) to detect and characterize genome-wide microsatellite loci with the following settings: Extend exact search; Repeat unit size range from 2 to 10; Maximum successive N's allowed in a repeat = 2; Recursion depth =5 ; Minimum and maximum percent perfection=80% and 100%, respectively (154). Finally, the average coverage and total number of comparative blastn hits for each *de novo* contig (NB1.0, NB1.1) also provided insight regarding unmasked repeats when cross referenced with the results of RepeatMasker (Tables S4, S13, S14) (170).

Following a two-stage RepeatMasker analysis (chicken + zebra finch repeat libraries), the masked NB1.1 scaffolds became the reference sequences used for SNP and indel prediction as previously described (153, 154, 274). After reference mapping all the trimmed sequence reads onto the double-masked NB1.1 assembly using the same assembly parameters described above, we used the CLC probabilistic variant detection algorithm (v6.0.4) to predict and estimate genome-wide variation (i.e., SNPs, indels) with the following settings: ignore nonspecific matches = yes; ignore broken read pairs = no;

minimum coverage = 10; variant probability ≥ 0.95; require variant in both forward and reverse reads = yes; maximum expected variants = 2; ignore quality scores = no. Histograms representing the NB1.1 coverage distribution of predicted genome-wide variants and their corresponding phred score distribution were produced using JMP Pro 10.0.1 (SAS Institute Inc., Cary, NC).

Effective Population Size Estimation

The bobwhite and scarlet macaw were chosen for comparison using PSMC (243) because they occupy opposing positions on the *r*-/*K*-selection continuum (57, 251, 252), with bobwhites being largely typical of an *r*-selected avian species, and the scarlet macaw clearly exhibiting characteristics of *K*-selection (2, 57, 251, 252). This allowed us to test the hypothesis that historic effective population size estimates for an *r*-selected avian species should theoretically exceed that of a *K*-selected avian species, and to compare the magnitude by which they differed. The input file for PSMC (243) was prepared according to the PSMC author's recommendations. For the bobwhite, variants with less than 46X coverage or more than 280X coverage were filtered from the diploid consensus. For the scarlet macaw, variants with less than 4X coverage or more than 26X coverage were filtered from the diploid consensus. Only NB1.1 and scarlet macaw (SMAC 1.1) (154) scaffolds aligning to autosomes were used. The maximum $2N_0$ coalescent time (parameter –t) was varied until at least 10 recombinations per atomic interval were observed. PSMC was run for 25 iterations, with –t10 –r5 –p "4+25*2+4+6" options used for the bobwhite and –t6 –r5 –p "4+25*2+4+6" used for the scarlet macaw. One hundred bootstraps were

used to calculate confidence intervals. We used the per-site pairwise sequence divergence to represent time and the scaled mutation rate to represent population size (243). To estimate generation time for the bobwhite, we evaluated long-term survivorship studies from across their U.S. range that did not rely on radio telemetry (30, 116, 275-278). Radio telemetry studies often greatly underestimate survivorship, so generation time based on such studies would also be underestimated (279). Bobwhite generation time (g) was estimated as: $g = a + [s / (1 - s)]$ (163, 280), where a = age of sexual maturity (~1 yr), and s = adult survival rate, as reported across the survivorship studies evaluated. We used the median generation time (1.22 yrs; range = 1.17 – 1.39 yrs) estimated across all studies for the bobwhite. At present, little is known about generation times in the scarlet macaw, with one source proposing a generation time of 12.7 years (http://www.birdlife.org/datazone/speciesfactsheet.php?id=1551&m=1). By considering an expected (s) of at least 90% across the scarlet macaw's range (i.e., in protected and unprotected regions), and (a) equivalent to 4 yrs, we estimated generation time for the scarlet macaw as approximately 13 yrs. Therefore, we used g = 12.7 in our PSMC analysis. Notably, our assumptions regarding s = 0.90 and a = 4.0 were both biologically feasible and reasonable, as evidenced by previous studies (281-283). Similar to recent PSMC analyses for the pig (*Sus scrofa*) genome (244), there are also no convincing data available regarding a different mutation rate in our birds (i.e., bobwhite, scarlet macaw) as compared to humans ($1.1–2.5\times10^{-8}$ mutations per generation) (271, 272). In fact, we initially estimated the substitution rate for the bobwhite and the scarlet macaw using autosomal genome alignment data and estimated divergence times as previously described

(163), but found that these estimates produced unreasonable PSMC results due to underestimation of the per-generation *de novo* mutation rate, as has been predicted by using the substitution rate (284). The most likely reasons for this are the relatively large estimated divergence times between the bobwhite and scarlet macaw as compared to other available, well annotated bird genomes (i.e., chicken, zebra finch, turkey), a very short generation interval for the bobwhite, a potential bias that is introduced by estimating the mutation rate via whole genome alignment (i.e., conserved regions align more stringently and more frequently), and the fact that the substitution rate only accounts for those mutations in lineages that persist in the face of drift and selection, which is not the same as the per-generation mutation rate observed from parent genome to offspring (284). For these reasons, we used two reasonable estimates for the mutation rate (i.e., 1.1×10^{-8} and the PSMC default value of 2.5×10^{-8} mutations per generation) to calibrate sequence divergence to years (243).

CHAPTER V

NORTHERN BOBWHITE (*COLINUS VIRGINIANUS*) MITOCHONDRIAL POPULATION GENOMICS REVEALS STRUCTURE, DIVERGENCE, AND EVIDENCE FOR HETEROPLASMY

Introduction

Subspecies may potentially play a vital role for conservation biologists and wildlife managers, who often seek to determine whether a species is demographically connected across its entire range, or if it is divided into subunits that are reflective of genetic and/or demographic structure (285). These distinctions are important for the accurate management of natural populations as well as for accurate taxonomic descriptions within a given taxon (285-292), since restricted gene flow promotes genetic sub-structuring among populations, thereby increasing genetic variance between populations (293). Two different concepts are a general basis for many management techniques, as follows 1) Ecological exchangeability, the idea that individuals can be moved between populations and can occupy the same ecological niche or selection régime because of shared fundamental adaptations of populations (i.e., similar life histories, ecological requirements, morphologies, and demographic characteristics) (289, 294, 295); and 2) Genetic exchangeability is the idea that individuals from different populations are genetically exchangeable if there is ample gene flow between populations (289), with levels of gene flow usually estimated via microsatellites, allozymes, nucleotides sequences (i.e., mtDNA,

nDNA, cpDNA, AFLPs, RAPDS), with each technique having its own strengths and weaknesses (296-298).

To date, more than 19 bobwhite subspecies have been named based on variation in size (decreasing from north to south) and male plumage (5, 10), with females displaying more similar plumage regardless of putative subspecies classification or geographic distribution (5, 10). Of these 19 putative subspecies, seven have been found in the United States (6, 285). Relevant to wild bobwhite populations in the southern U.S. and northern Mexico, four putative subspecies have been recognized west of the Mississippi River, which include the eastern (*C. v. virginianus*), plains (*C. v. taylori*), Texas (*C. v. texanus*), and masked bobwhite (*C. v. ridgwayi*) (9, 10), which is phenotypically and geographically distinct (i.e., black head; Sonora, Mexico) (9, 10).

The desire to mitigate U.S. bobwhite population declines has prompted both the translocation of wild bobwhites to fragmented regions of their historic range as well as attempts at restocking or population supplementation using pen-raised stock; with neither of these approaches being reported to be highly successful in regions where modern abundance is low (97-102). At present, a need exists to examine the genetic relationships and overall levels of divergence within and between putative bobwhite subspecies as well as their extant U.S. populations; to enable informed management and restoration efforts.

Historically, avian population studies have utilized the D-Loop to elucidate population structure (299, 300). The control region (i.e., D-Loop) surrounds the origin of replication of the mtDNA molecule (301) and is considered to be the most variable portion of the mitochondrial genome (mitogenome) (301). The D-Loop itself is a noncoding

region which suggests that there is a lack of functional constraint(s) thus explaining its hyper variability (301-303). The utility of the D-Loop for revealing previously undetected genetic structure within and among closely related avian taxa has been established (299), with putative subspecies detected in dunlins (304, 305), grey-crowned babblers (293), and lesser snow geese (306).

To date, the mitogenome has been an important molecule used in many population studies. One primary reason for this has been the ease of its use, and more specifically, the ability to easily isolate mtDNA from a variety of tissue types, the simplicity of mitogenome architecture, the general lack of recombination, relatively high levels of nucleotide diversity, and maternal inheritance (i.e., haploid) (301, 304, 305, 307). Moreover, mitogenomes contain a combination of both slowly evolving 'highly conserved' regions and rapidly evolving 'highly variable' regions (308-312), which has made mtDNA suitable for species and population level studies (301). However these variable evolutionary rates may reflect different evolutionary histories, depending on what regions are targeted for study (296), which makes the use of multiple genetic markers necessary to accurately resolve species-level evolutionary and taxonomic relationships (307), but may also require the use of nuclear markers for corroborating inferences.

A recent bobwhite mitochondrial DNA (mtDNA) study reported a general lack of distinct phylogeographic structure, evidence for demographic expansion following the Pleistocene, and an apparent discordance between patterns of mtDNA diversity and subspecies designations for the four putative subspecies west of the Mississippi River; with all inferences based on the analysis of a 353 bp fragment from the mitochondrial control

region (9). Given the availability of a draft nuclear and mitochondrial genome assembly for the bobwhite (313), large-scale genetic studies are both possible and warranted, especially considering the apparent decline of wild bobwhite populations across the majority of their historic U.S. range (73, 74, 77-80). We generated complete mitochondrial genome sequences for bobwhites sampled from six discrete ecoregions across Texas and Oklahoma (USA), including representative samples from two putative bobwhite subspecies (*C. v. texanus*; *C. v. taylori*). Thereafter, we evaluated whether small mitochondrial fragments (i.e., partial or complete D-loop) could accurately resolve and predict the true haplotype structure and relationships among our samples, as compared to using complete mitogenome sequences to perform the same analyses. We further tested this same hypothesis with respect to accurately inferring historical patterns of demography, signatures of population substructure, and whether or not partial or complete mitogenomic data would support the presence of two or more putative bobwhite subspecies. The results of this study provide new insights regarding the demographic history and diversity of bobwhite maternal lineages west of the Mississippi River, but also clearly underscore the need for large-scale genomic studies in declining wildlife species.

Results and Discussion

Bobwhite Mitogenome Sequencing, Reference Mapping, and Variant Detection

Herein, we generated complete mitogenome sequences for 51 bobwhites representing two U.S. states (TX, OK) and 6 discrete ecoregions [see Appendix B] using standard Illumina paired-end (PE) sequencing technologies (i.e., TruSeq PE 2 x 100 bp;

Illumina HiSeq2500; see Methods). Thereafter, we used these sequences and one additional bobwhite mitogenome (GenBank KJ914548.1) obtained from a phylogenetic study of the Odontophoridae (New World quail) (314) to predict single nucleotide variants (SNVs) and insertion-deletion mutations (indels) via reference mapping and alignment to an updated bobwhite mitogenome reference sequence (313) (n = 53 total bobwhite mitogenomes; see Methods). Using a probabilistic variant detection algorithm previously described (313) (see Methods), we predicted 344 segregating sites corresponding to 347 total mutations (n = 338 SNVs, 8 indels, and 1 multi-nucleotide variant, MNV), which included 49 putative nonsynonymous SNVs distributed across 12 protein coding genes. The majority of the nonsynonymous SNVs (i.e., 80%) were predicted at relatively low frequencies (i.e., < 0.10), and *ND4L* was the only mitochondrial protein coding gene for which zero nonsynonymous variation was predicted. However, eight of the nonsynonymous SNVs were predicted at moderate frequencies (i.e., > 0.10) in our samples, with corresponding amino acid replacements predicted in five mitochondrial protein coding genes (i.e., *CYTB*, *COX1*, *ATP6*, *COX3*, *ND5*). Similar to several previous avian and reptile studies, we also found an unambiguous *ND3* single nucleotide insertion (i.e., frameshift) in all bobwhite mitogenome sequences that were generated during this study [for review see 314-322]. Moreover, we also compared all predicted variants and their proximal flanking sequences to the known galliform nuclear mitochondrial sequences (numts) previously described (323), which included those identified in the first-generation draft genome assembly for the bobwhite (313). With the exception of *ND3* (314-322), no indels or premature stop codons were observed in any bobwhite mitochondrial protein

coding genes. However, two discrete SNVs were observed which could not be unequivocally excluded as potential numts, and therefore, we excluded these from all subsequent analyses.

Bobwhite Mitogenome Heteroplasmy

The ability to generate bobwhite mitogenomes with deep coverage using Illumina PE sequencing technologies provided an opportunity to investigate the potential for heteroplasmy (324, 325), which has been reported in several avian species (321, 326-331), with one study indicating that paternal leakage may be a key factor in the emergence of some avian heteroplasmies (327). Microheteroplasmy, which is defined by rare (i.e., independent) mutations found among 1-2% of all intra-individual mitogenomes, is common among adult humans, and has led some researchers to postulate whether this mutational burden may be linked to aging as well as age-related diseases (332-335). However, microheteroplasmy can be differentiated from gross heteroplasmy by the presence of moderate to high frequency (i.e., common) mutations observed among the mitogenomes recovered from a single individual and/or a discrete tissue (332-335). We detected evidence for gross mitochondrial heteroplasmy in 13 of the 51 surveyed bobwhites (i.e., 25%), which is similar to the heteroplasmy rates (i.e., 24%) reported for a survey of five human populations (324), and those reported for the crested ibis (*Nipponia Nippon*) (i.e., 22%) (331). Specifically, in 13 of the 51 surveyed bobwhites, we identified 16 moderate to high frequency heteroplasmies (i.e., heterozygous mitochondrial sites) with minor allele frequencies ranging from 22% to 46.5%. All 16 detected heteroplasmies

involved single nucleotide variants (SNVs) possessing average quality scores > 32, with 14 of the 16 (87.5%) observed as singletons among our population samples. Ten of the 13 bobwhites were predicted to possess only one heteroplasmic SNV (i.e., two unambiguous mtDNA haplotypes), whereas the other three heteroplasmies involved two unambiguous (n = 3 bobwhites) intra-individual heteroplasmic SNVs. Two of the 16 detected heteroplasmies (SNV sites 2216 and 2418) were also individually observed as homozygous SNVs (i.e., on different mtDNA haplotypes) in a second bobwhite sequenced during our population survey. The distribution of the 16 heteroplasmic sites included both coding and noncoding regions (i.e., tRNA-Val, D-Loop, *12S*, *COX2*, *ATP6*, *CYTB*, *ND1*, *COX1*, and *COX3*), with 8 SNVs that were predicted to encode amino acid substitutions. As previously described, four plausible biological mechanisms may facilitate heteroplasmy including: 1) Paternal leakage; 2) Maternal transmission/inheritance of heteroplasmic variants; 3) *De novo* mutations that occur during embryonic development; and 4) Somatic aging, with age-related accumulation of heteroplasmic variants (324, 325, 327, 332, 334, 335). In the absence bobwhite samples of known pedigree, we could not unequivocally attribute the observed heteroplasmies to either paternal leakage or maternal inheritance of heteroplasmic sites. However, an evaluation of all the bobwhite mtDNA haplotypes generated in this study provides sufficient information (i.e., via variable sites) to predict the expected signatures of DNA contamination (i.e., the expected heterozygous mtDNA sites resulting from mixed samples). No evidence of contamination was observed. We also examined the distribution of ages among all of the heteroplasmic bobwhites observed in this study, and found nearly equal proportions of both juveniles and adults, indicating that

somatic aging is unlikely to explain the observed heteroplasmies. Additionally, heteroplasmy was detected for bobwhites representing both putative subspecies (5, 9, 10) (*C. v. taylori*, n = 7; *C. v. texanus*, n = 6) sampled from five ecoregions and two U.S. states (i.e., Southwestern Tablelands of TX and OK, Western Gulf Coastal Plain of TX, Central Great Plains of TX and OK, Southern Texas Plains of TX, High Plains of TX and OK). Notably, most instances of bobwhite heteroplasmy detected in this study (i.e., 10 / 13 = 76.9%) relates to the presence of two intra-individual mtDNA haplotypes that differ by one SNV, which most likely arose via paternal leakage, maternal transmission, or *de novo* mutation during embryonic development. Future studies that include larger sample sizes of known pedigree are needed to deduce the biological mechanism(s) underlying instances of gross heteroplasmy in the bobwhite.

Bobwhite Population Structure, Phylogeography, and Historical Demography

Herein, we conducted a series of comparative analyses to determine whether similar population inferences or conclusions could be deduced from partial (i.e., partial or complete D-loop) and complete mitogenome nucleotide sequence data for 53 bobwhites (n = 6 ecoregions across TX and OK). As expected, haplotype diversity increases with the inclusion of increasing levels of mitogenomic sequence data, and nucleotide diversity decreases, the latter being due to the fact that nucleotide diversity is directly impacted by localized hyper-variability within short fragments of the mitochondria that are commonly targeted for population analyses (i.e., partial or complete D-Loop) [for review see 9, 336-340]. Median joining haplotype networks (341) constructed for partial bobwhite

mitogenome sequences demonstrated an overt lack of resolution for drawing phylogenetic or population inferences in the bobwhite, as compared to networks constructed using complete mitogenome sequence data (Fig. 5, Fig. 6). Moreover, for analyses which utilized partial sequences, the reduction in mitochondrial genomic information content was observed to encourage spurious inferences in our samples (Table 6, Fig. 5, Fig. 6). For example, the total number of unique mitochondrial haplotypes and haplotype diversity were highly underestimated (i.e., collapsed) when partial mitogenome sequences were utilized, and therefore, some bobwhites wrongly appear to possess identical mitochondrial haplotypes (Table 6, Fig. 5, Fig. 6). This problem should be expected in many studies which utilize small mitogenome fragments, rather than complete mitogenome sequences. A comparative summary of all bobwhite mitochondrial analyses of diversity are presented in Table 6. Moreover, the true degree of mitogenome divergence and population structure among our sampled bobwhites was not detectable when popular mitogenome fragments (i.e., partial or complete D-Loop) [for review see 9, 336-340] were analyzed (Fig. 5, Fig. 6). Nevertheless, similar to a previous bobwhite mitochondrial study (9), we did not observe strong phylogeographical clustering among the six surveyed U.S. Environmental Protection Agency (EPA) level III ecoregions (http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm; Fig. 5). However, it should be noted that among the two discrete mitogenome haplotype groups detected (Fig. 5, Fig. 6), many of the diverged individuals (i.e., Group 2, Fig. 6) originated from one ecoregion (n = 8 / 17, or 47%, South Texas Plains). The precise origin of this previously undetected

diverged lineage (9), which represents approximately 25% of the total bobwhites surveyed in this study, is currently unknown.

Considering partial or complete bobwhite mitogenome sequences, we observed little support for the previously described geographic subspecies designations (5, 9, 10) across the six investigated ecoregions. However, complete mitogenome sequence analyses did reveal a modern bobwhite population structure that may potentially be comprised of at least two putative subspecies (F_{ST} = 0.849; $P < 0.05$; Fig. 5, Fig. 6); with the divergence between these two groups almost exclusively observed for a subset of bobwhites geographically classified as *C. v. texanus* (5, 9, 10) (Fig. 5). Additional analyses further demonstrated statistically significant mitogenome differentiation and population subdivision between the two groups (i.e., K_S, K_S^*, Z, Z^*, $P < 0.001$ via permutation) (342). The average number of nucleotide substitutions per site between the two lineages (Group 1 versus Group 2; Fig. 5, Fig. 6) was 0.00731 (d_{XY}) (343), indicating that the average percent divergence was less than 1% (i.e., 0.7%). Collectively, 103 mitogenome mutations defined the split between the two bobwhite lineages within a median joining haplotype network (i.e., 101 SNVs, 2 Indels; Fig. 5, Fig. 6). Examination of all 103 network torso mutations revealed eight SNVs that were predicted to cause amino acid replacements, and 50% of these localized to *ND5*. No heteroplasmic variable sites were present in the network torso. Similar to our median joining haplotype networks, complete mitogenome divergence was also detected and visualized via mismatch distribution, where a bimodal distribution becomes overtly apparent with the inclusion of increasing levels of mitogenome sequence data (Fig. 7). This bimodal distribution is in conflict with a previous bobwhite study that

reported a unimodal mismatch distribution (i.e., based on a 353 bp mitogenome fragment), and corresponding inference suggesting a recent, rapid demographic expansion (9). As shown in Figure 7, our analysis of the same 353 bp mitogenomic region produced a mismatch distribution that was strikingly similar to that of Williford and colleagues (9). However, the true mismatch distribution corresponding to the extant maternal lineages sampled during this study was very poorly estimated when only 353 bp (i.e., partial D-Loop) were analyzed, indicating that more sequence data is necessary to correctly infer aspects of bobwhite historical demography and/or population substructure (Fig. 7). Bimodal or multimodal mitochondrial mismatch distributions have not been uniformly interpreted in the literature; with some authors suggesting that these distributions reflect stable, stationary populations (i.e., post expansion) with or without spatial structuring (344-346), populations that are expanding spatially via few outward migrants per generation (347), or populations with tangible substructure and/or mutation rate heterogeneity (i.e., even while experiencing demographic expansions) (348, 349). Therefore, the true biological origin(s) of any bimodal mismatch distribution may be complex. For example, factors such as biogeographical barriers (346), survival of some divergent lineages from a pre-expansion to a post-expansion population (350), the occurrence of historic population admixture (351), and even hybridization (352) have all been noted as likely origins. In this study, a comparison of all bobwhites geographically classified as either *C. v. texanus* or *C. v. taylori* (5, 9, 10) produced F_{ST} values that were statistically significant ($P < 0.05$; Table 7), which is concordant with a recent study (9), but notably, these F_{ST} values are far smaller than those obtained for a comparison of the two bobwhite mitogenome lineages

elucidated by median joining haplotype networks (See Fig. 5C, Fig. 6C; Group 1 vs Group 2). Although little evidence of strong phylogeographic clustering and corresponding subspecies distributions were observed in this study, and/or during a previous study (9), the F_{ST} values obtained via comparison of bobwhites that were taxonomically classified based on geographic subspecies designations (5, 9, 10) (Table 7; *C. v. texanus* versus *C. v. taylori*) suggests that perhaps a more pronounced substructure may have existed, historically, but has since been diluted. For example, the composition of wild bobwhite populations may potentially be affected by an agricultural practice that involves the introduction of pen-reared lineages for restocking or supplementation (97, 100-102). Herein we show that at least some pen-raised lineages are indistinguishable at the mitogenome sequence level from that of wild bobwhites (Fig. 5C, Fig. 6C), while others have been distinguished using nuclear microsatellite loci (102). Interestingly, annual survival and breeding of pen-reared bobwhites (pen-reared x pen-reared; pen-reared x wild) following release into suitable habitats has often been considered low (97, 100, 101). However, some previous studies actually demonstrate either tangible annual survival rates (97, 102), and/or apparent reproduction (pen-reared x pen-reared; pen-reared x wild) (102). Moreover, a study which previously concluded low post-release survival for pen-reared bobwhites, and subsequently discouraged their use for restocking, actually shows a basal survival rate that is greater than 20% across the entire observation period (i.e., 22 weeks) at one of the two study sites evaluated (97). Therefore, it is apparent that some proportion of pen-reared bobwhites may successfully integrate into some wild populations (97, 102). In this study, we produced complete mitogenome sequences for one hunter harvested pen-

reared bobwhite (i.e., marked by a leg-band), and two bobwhites that were hunter harvested in a pasture routinely used for pen-reared releases. Complete mitogenome haplotypes corresponding to these three bobwhites were observed in both of the two divergent mitogenome clusters (n = 2 bobwhites in Group 1; n = 1 bobwhite in Group 2; Fig. 5C, Fig. 6C), with the known pen-reared bobwhite possessing a haplotype that was more closely related to the most common wild bobwhite mitogenome sequences (i.e., Group 1; Fig. 5C, Fig. 6C).

Since the majority of the diverged mitogenomes (i.e., Group 2, Fig. 5C, Fig. 6C) were recovered from bobwhites that occupied overlapping ranges with the scaled quail (*Callipepla squamata*; also known as the blue quail), we investigated whether hybridization might be explanatory for the observed divergence. This hypothesis was predicated on previous observations that bobwhites and scaled quail may hybridize, both in the wild and in captivity (353, 354). To address this question, we used standard Illumina PE sequencing technologies to produce a complete mitogenome sequence (n = 16,701 bp; GenBank Accession KT722338; see Methods) for a hunter harvested scaled quail that was obtained from the same ranch where multiple diverged bobwhites were sampled (Group 2, Fig. 5C, Fig. 6C). Comparison of the scaled quail and bobwhite reference mitogenome revealed 1,215 mutational differences (n = 1,191 SNVs; n = 24 Indels), or greater than 7% divergence, indicating that hybridization between these two species is not explanatory for the two diverged bobwhite mitogenome haplotype groups (Fig. 5C, Fig. 6C). Moreover, a maximum likelihood-based phylogeny constructed with expanded taxon sampling demonstrated that that the scaled quail is more closely related to the bobwhite than the

tawny-faced quail (*Rhynchortyx cinctus*) (355), and that both bobwhite haplotype groups were equidistant from the scaled quail (Fig. 8). Likewise, using the scaled quail as an outgroup Tajima's relative rate test (356) revealed no significant rate heterogeneity between the two bobwhite lineages. These results are interesting because they suggest that neither bobwhite mitogenome group is more ancestral (or more derived) than the other, which supports the hypothesis that divergent maternal lineages have survived from a pre-expansion to a post-expansion population (350). To further address this question, we examined the individual mismatch distributions for each bobwhite mitogenome group that was identified (i.e., Group 1, Group 2; Fig. 5C, Fig. 6C). Both groups individually fit a demographic model of population growth-decline better than a model which assumed a stable, constant population size (357-359) (Fig. 9). This result was robust to using either the mismatch distribution (i.e., pairwise number of differences) and/or the site frequency spectrum (i.e., segregating sites; not shown), and is generally concordant with previous reports of range-wide declines for the bobwhite (Fig. 9) (73-79). One final inference that could be drawn from our maximum likelihood-based phylogenies pertained to the likely origin of most bobwhite heteroplasmies. Specifically, in all instances of heteroplasmy, the two intra-individual mitogenome haplotypes were observed as sister taxa within the phylogenetic tree, thereby suggesting that either maternal transmission and/or developmental *de novo* mutation(s) were the most likely origin(s).

Application of Tajima's and Fu's test (i.e., D, F_S) (360, 361) to each bobwhite mitogenome group (Fig. 5C, Fig. 6C) revealed negative test statistics for both clusters (Group 1, n = 49, D = -1.83, F_S = -25.93; Group 2, n = 17, D = -1.25, F_S = -4.24).

However, these tests were only statistically significant for Group 1 (i.e., via beta distribution and coalescent simulations; See Methods). Specifically, this result occurs due to an excess of rare variants and rare haplotypes in Group 1, which is consistent with new (i.e., young) mutations resulting from demographic expansion and/or natural selection (i.e., positive or purifying). Application of a multi-locus McDonald-Kreitman test (362-365) (MKT, including all 13 mitochondrial protein coding genes) to evaluate the possibility for functional divergence among the two bobwhite groups (Fig. 5C, Fig. 6C) revealed some putative evidence for purifying selection ($P < 0.05$; Overall mean proportion of adaptive substitutions (α) = -1.159), and no evidence for functional divergence via positive selection. More specifically, it should be noted that deleterious mutations rarely become fixed in a population (362), and in this case, a significant multi-locus MKT results from the high proportion of polymorphism (including singleton heteroplasmic SNVs) as compared to the very low proportion of fixed (i.e., diverged) nucleotide sites between the two bobwhite mitogenome lineages detected (Fig. 5C, Fig. 6C). This interpretation can be further evidenced by removing all heteroplasmic SNV haplotypes, the majority of which were singletons, and again computing the multi-locus MKT for the two bobwhite lineages, which revealed no evidence for significant purifying selection ($P > 0.05$; (α) = -0.774). Likewise, standard MKT's (362, 363) for all individual mitochondrial protein coding genes also revealed no evidence for positive selection and functional divergence ($P > 0.05$) regardless of the inclusion or exclusion of heteroplasmic (i.e., low frequency) mitochondrial haplotypes. Fine scale analyses conducted using Tajima's test (D) via sliding window (i.e., 100 bp window, 25 bp step) also failed to produce any statistically

significant evidence for natural selection within the mitochondrial protein coding genes for members of the two bobwhite mitogenome haplotype clusters (i.e., individual analyses for Group 1; Group 2). Similarly, application of Tajima's and Fu's Tests (*D*, F_S) (360, 361) to the pooled set of all bobwhite mitogenome haplotypes (Group 1 + Group 2) also provided little support for strong selection, and less demographic insight relative to individual analyses carried out for members of each haplotype group (i.e., $D = -0.59$, $P > 0.05$; $F_S = -16.24$, $P < 0.01$ by coalescent simulations). Significantly negative values observed for *D* and F_S (360, 361) in relation to the Group 1 bobwhites (Fig. 5C, Fig. 6C) most likely reflects signatures of demographic expansion in that lineage and/or purifying selection (i.e., perhaps detectable via larger Group 1 sample size). Application of Tajima's test (*D*) to the pooled bobwhite sample (Group 1 + Group 2) using the sliding window method (i.e., 100 bp window, 25 bp step) revealed one mitogenome window located in the *ND5* gene (C-terminal region of NADH5; Pfam NADH5_C Domain) which departed from the neutral expectation ($D = -1.83$, $P < 0.05$). This result was driven by the occurrence of four singleton SNVs within the 100 bp window (n = 3 nonsynonymous; n = 1 synonymous). The distribution of these SNVs included both bobwhite mitogenome lineages (Fig. 5, Fig. 6) identified by network analyses (n = 2 nonsynonymous and 1 synonymous in Group 1; n = 1 nonsynonymous in Group 2). Therefore, while the individual MKT (362, 363) for *ND5* of bobwhite Group 1 versus Group 2 (Fig. 5, Fig. 6) included four nonsynonymous substitutions that were fixed differences between the two mitogenome groups, that test was not statistically significant. Therefore, the collective results from our analyses of these data (i.e., pooled and by individual groups) are most concordant with the segregation of some

slightly deleterious nonsynonymous mutations, as further evidenced by the negative overall mean proportion of adaptive substitutions, and the absence of statistical support for positive selection or functional divergence between the two bobwhite mitogenome lineages (362-365).

Table 6. Bobwhite Mitochondrial Analyses of Diversity.

Summary Data*	**Partial D-Loop**	**Complete D-Loop**	**Complete Mitogenome**
Sample size (haplotypes)	54	55	66
Size of analyzed region (bp)	353 bp	1152 bp	16,702 bp
Total variable sites	19	31	335
Total number of mutations	20	33	338
Total unique haplotypes	22	34	62
Haplotype diversity (Hd)	0.860	0.966	0.998
Nucleotide diversity (π)	0.00868	0.00435	0.00354

* Includes heteroplasmic minor allele haplotypes, excluding gaps.

Table 7. Pairwise F_{ST} Values between Geographically Designated Bobwhite Subspecies.

	Partial D-Loop (353 bp)		**Complete D-Loop (1152 bp)**		**Mitogenome (16,709 bp)**	
	C. v. texanus	*C. v. taylori*	*C. v. texanus*	*C. v. taylori*	*C. v. texanus*	*C. v. taylori*
C. v. texanus		≤ 0.0001		≤ 0.0001		≤0.0001
C. v. taylori	0.25407*		0.18956*		0.31271*	

* Significant ($P < 0.05$) F_{ST} values (below diagonal, with standard errors above diagonal)

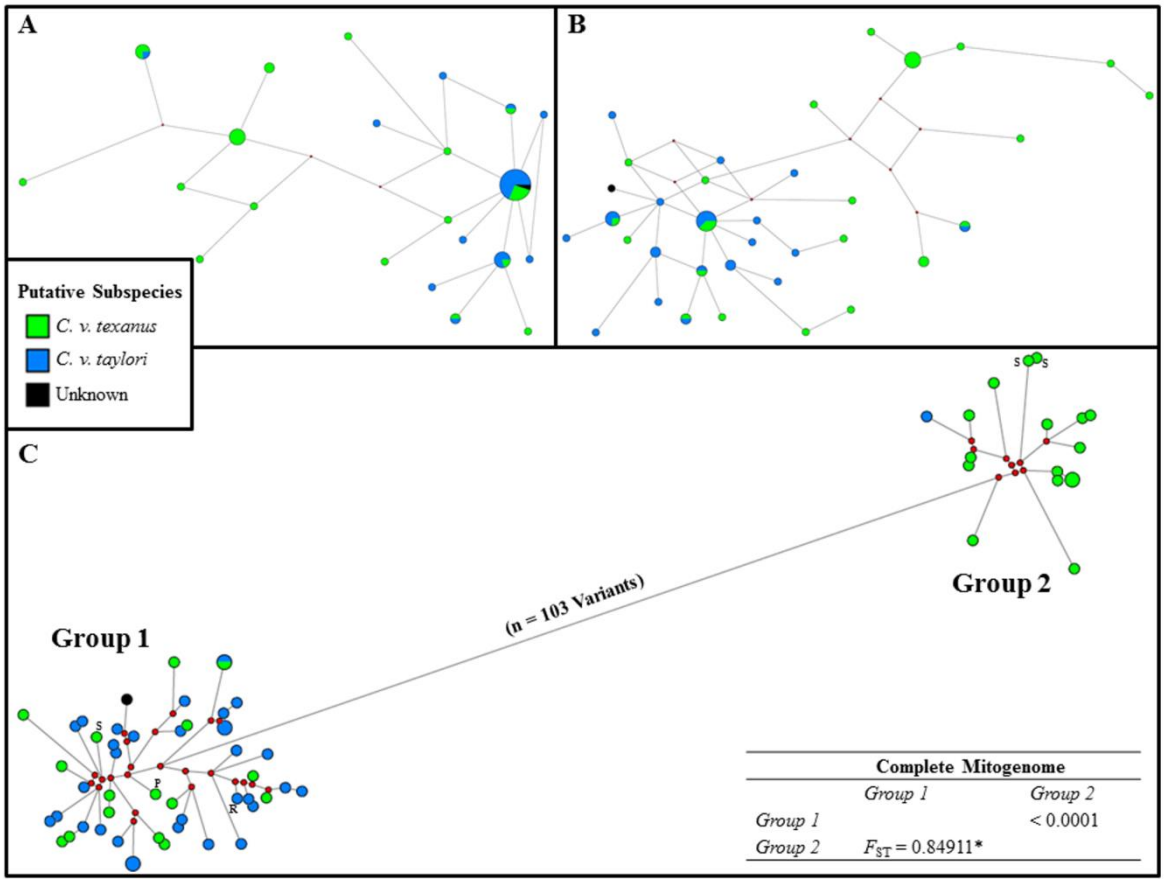


Figure 5. Median Joining (MJ) Haplotype Networks (341) Constructed for Partial and Complete Bobwhite Mitogenome Sequences, with Heteroplasmic Minor Allele Haplotypes, and Color-coded Geographic Subspecies Designations (5, 9, 10). (A) MJ haplotype network for 353 bp of the mitochondrial D-Loop (9) (n = 54, including 1 heteroplasmic minor allele haplotype). (B) MJ haplotype network for the complete D-Loop (1152 bp; n = 55, including 2 heteroplasmic minor allele haplotypes). (C) MJ haplotype network for the complete mitogenome (16,709 bp including gaps; n = 66, including 13 heteroplasmic minor allele haplotypes). Default weights for SNPs and indels were used (10 and 20, respectively), with node sizes proportional to haplotype frequency, and branch lengths drawn to scale. Red dots indicate median vectors. The complete mitogenome haplotypes were observed to form two divergent clusters (i.e., Group 1, Group 2; n = 103 variants). Pairwise F_{ST} values (below diagonal) with standard errors (above diagonal) were computed to assess genetic differentiation between the two clusters, with the asterisk (*) indicating a significant F_{ST} value ($P < 0.05$). Figure 5C includes three complete mitogenome haplotypes for bobwhites lawfully harvested from active surrogating pastures (i.e., pen release sites = S), and one haplotype from a lawfully harvested pen-released bobwhite (P). (R) Indicates the reference mitogenome (313).

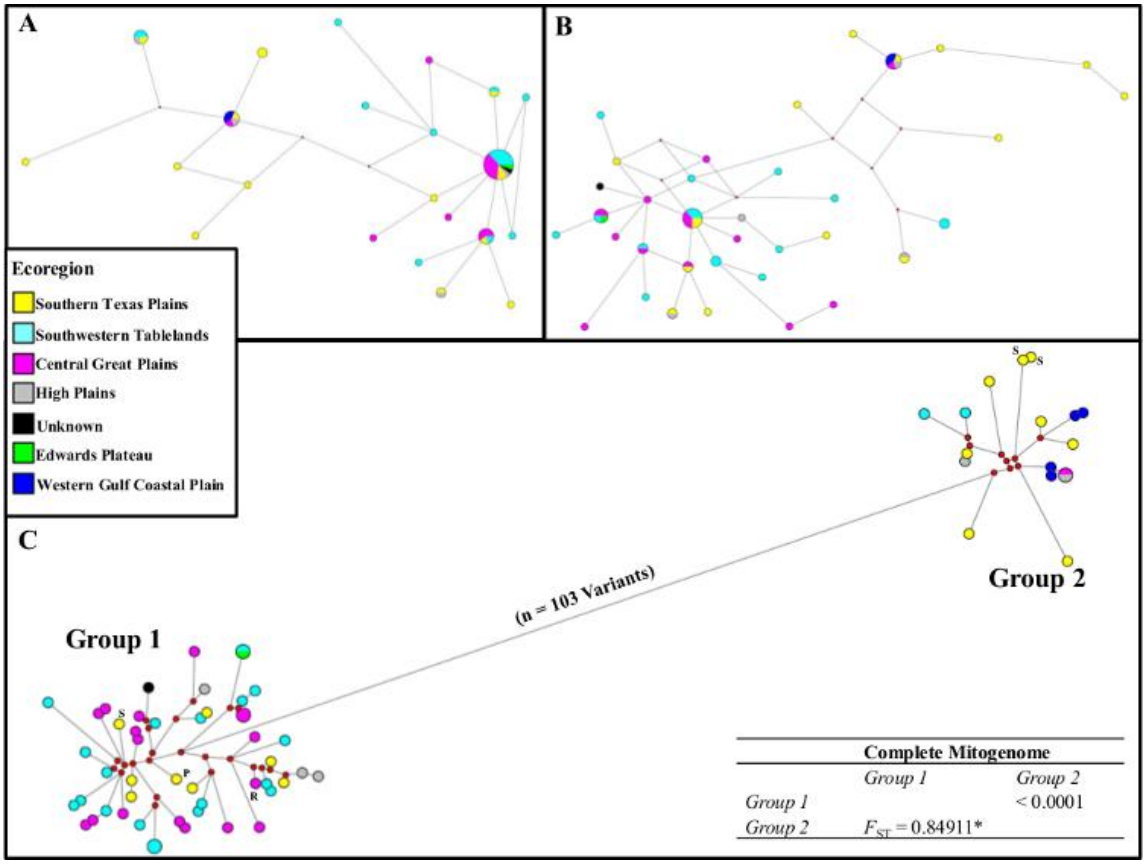


Figure 6. Median Joining (MJ) Haplotype Networks (341) Constructed for Partial and Complete Bobwhite Mitogenome Sequences, with Heteroplasmic Minor Allele Haplotypes, and Color-coded Assignments to U.S. Environmental Protection Agency Level III Ecoregions (http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). (A) MJ haplotype network for 353 bp of the mitochondrial D-Loop (9) (n = 54, including 1 heteroplasmic minor allele haplotype). (B) MJ haplotype network for the complete D-Loop (1152 bp; n = 55, including 2 heteroplasmic minor allele haplotypes). (C) MJ haplotype network for the complete mitogenome (16,709 bp including gaps; n = 66, including 13 heteroplasmic minor allele haplotypes). Default weights for SNPs and indels were used (10 and 20, respectively), with node sizes proportional to haplotype frequency, and branch lengths drawn to scale. Red dots indicate median vectors. The complete mitogenome haplotypes were observed to form two divergent clusters (i.e., Group 1, Group 2; n = 103 variants). Pairwise F_{ST} values (below diagonal) with standard errors (above diagonal) were computed to assess genetic differentiation between the two clusters, with the asterisk (*) indicating a significant F_{ST} value ($P < 0.05$). Figure 6C includes three complete mitogenome haplotypes for bobwhites harvested from active surrogating pastures (i.e., pen release sites = S), and one haplotype from a lawfully harvested pen-released bobwhite (P). (R) Indicates the reference mitogenome (313).

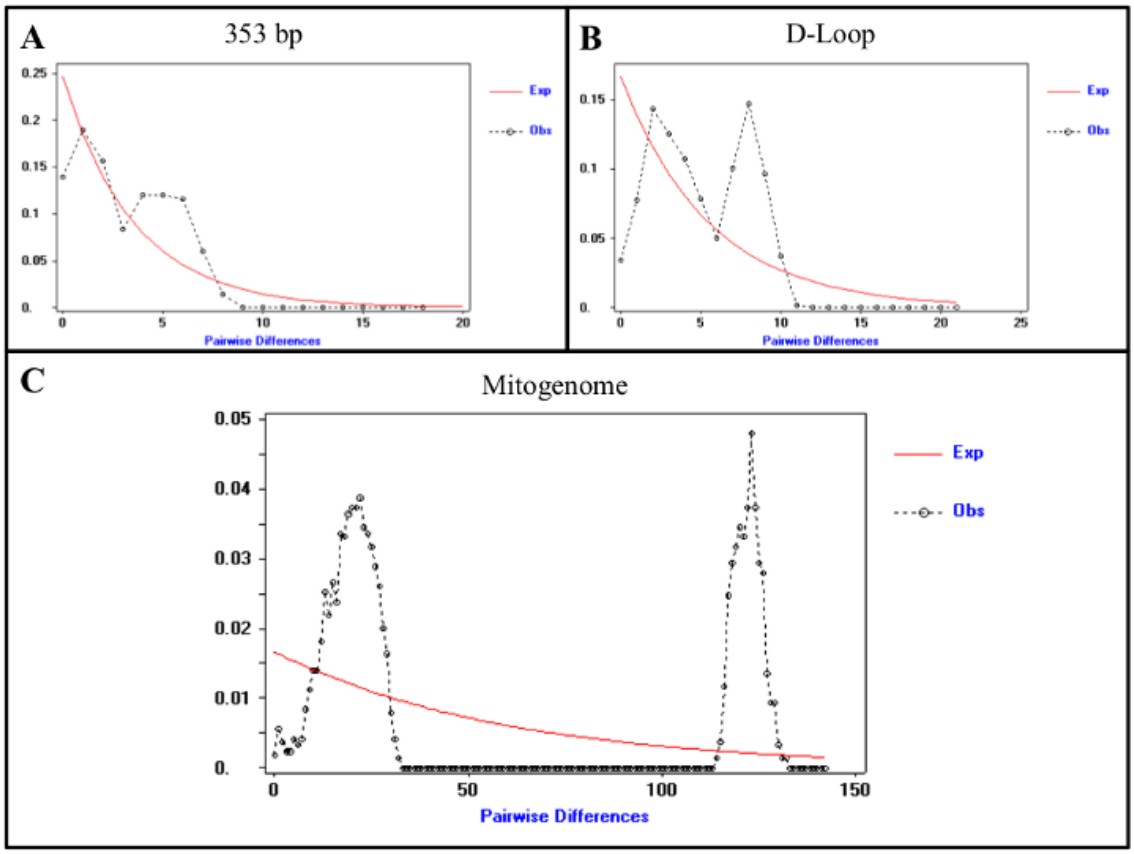


Figure 7. Bobwhite Mismatch Distributions. (A) 353 bp of the mitochondrial D-Loop (9) (n = 54, including 1 heteroplasmic minor allele haplotype). (B) Complete D-Loop (1,151 bp excluding gaps; n = 55, including 2 heteroplasmic minor allele haplotypes). (C) Complete mitogenome (16,698 bp excluding gaps; n = 66, including 13 heteroplasmic minor allele haplotypes). The x-axis represents the number of pairwise differences (mismatches) and the y-axis represents the frequency of these differences. The observed mismatch distribution (dashed line) is compared to the expected distribution (red line) for a stable population (i.e., constant population size).

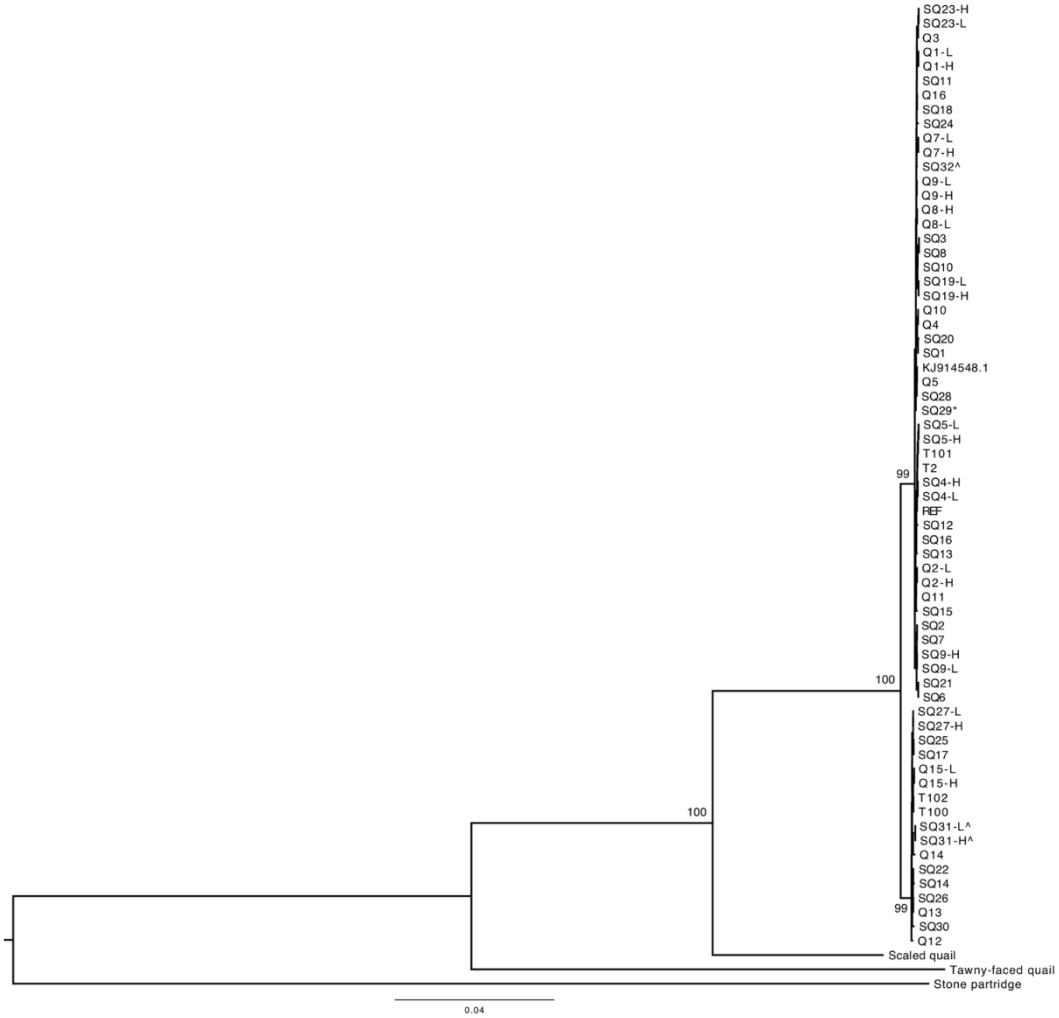


Figure 8. Maximum Likelihood-based Phylogeny Constructed with Expanded Taxon Sampling. Phylogeny of all bobwhite mitogenomes (n = 66, including 13 heteroplasmic minor allele haplotypes) in conjunction with mitogenomes for the scaled quail (*Callipepla squamata*; GenBank Accession KT722338), tawny-faced quail (*Rhynchortyx cinctus*; GenBank Accession KJ914547.1) (314), and stone partridge (*Ptilopachus petrosus*; GenBank Accession KJ914543.1) (314). The asterisk (*) denotes a pen-released (n = 1) origin, and '^' denotes from active surrogating pastures (i.e., pen release sites mitogenomes belonging to birds that were obtained from land used for pen-reared releases (n = 3). Terminal taxa noted with "H" and "L" refers to the high frequency (i.e., major allele) and low frequency (i.e., minor allele) heteroplasmic haplotypes, respectively. Individual bobwhites are labeled with laboratory identifiers (Q, SQ, T), with REF indicating the bobwhite reference sequence (GenBank Accession AWGT00000000.1), and KJ914548.1 indicating a bobwhite GenBank Accession (314) included in our analyses. The maximum likelihood phylogeny was constructed with RAxML 7.2.8 (366) using a GTR+Γ model of sequence evolution, with bootstrap support values based on 1,000 pseudoreplicates.

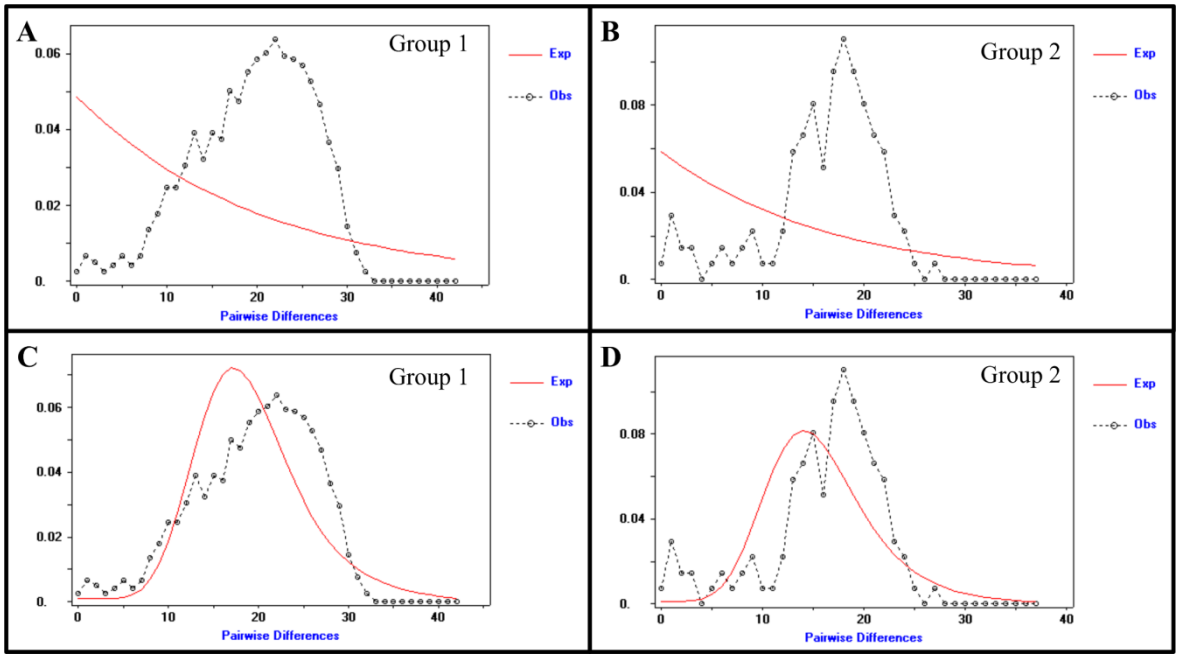


Figure 9. Bobwhite Historical Demography Inferred from Complete Mitogenome Sequences. (A) The observed mismatch distribution (dashed line) for bobwhite Group 1 (n = 49) as compared to the expected distribution (red line) for a stable population (i.e., constant population size). (B) The observed mismatch distribution (dashed line) for bobwhite Group 2 (n = 17) as compared to the expected distribution (red line) for a stable population (i.e., constant population size). (C) The observed mismatch distribution (dashed line) for bobwhite Group 1 (n = 49) as compared to the expected distribution (red line) for a growth-decline model. (D) The observed mismatch distribution (dashed line) for bobwhite Group 2 (n = 17) as compared to the expected distribution (red line) for a growth-decline model.

Methods

Bobwhite and Scaled Quail Sampling, Taxonomy, and Isolation of Genomic DNA

Two sources of bobwhite quail (n = 25 females; n = 26 males) were utilized for DNA isolation in the present study, including lawfully harvested wild bobwhites for which ethical clearance is not applicable, and those collected via trapping, where ethical clearance is required. Bobwhites obtained via trapping (n = 27) were collected during two-week periods of August and October (2012 and 2013) using milo-baited funnel traps on private ranches and public wildlife management areas located in the Central Great Plains, Edwards Plateau, High Plains, and Southwestern Tablelands ecoregions of Texas and Oklahoma. Birds were collected under authorization of a Texas Parks and Wildlife permit (SPR-1098-984; Austin, TX, USA) and via Institutional Animal Use Protocols from both Texas Tech University (IACUC 11049-07; Lubbock, TX, USA) and Texas A&M University (IACUC 2011-93; College Station, TX, USA). Skeletal muscle samples (i.e., from one or both legs) were obtained from bobwhites that were lawfully harvested (n = 24) on private ranches in the Central Great Plains, Southwestern Tablelands, Southern Texas Plains, and Western Gulf Coast Plain ecoregions of TX (USA) [see Appendix B]. Likewise, skeletal muscle samples from the legs of one lawfully harvested scaled quail were also obtained from one of the same private ranches in the Southern Texas Plains ecoregion. Bobwhite ecoregion assignments followed the U.S. Environmental Protection Agency (EPA) level III ecoregion maps (http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). Putative subspecies designations for all bobwhites included in this study (i.e., *C. v. texanus*; *C. v. taylori*)

followed geographic designations described west of the Mississippi River (USA) (5, 9, 10). Genomic DNA was isolated from skeletal muscle derived from the legs of bobwhites (n = 52) and one scaled quail (n = 1) using the MasterPure DNA Purification Kit, according to the manufacturer's recommendations (Epicentre Biotechnologies Inc., Madison, WI) (313). The presence of high molecular weight genomic DNA was assessed and confirmed by agarose gel electrophoresis, with quantitation via Nano Drop 1000 (NanoDrop Technologies Inc., Wilmington, DE), and by evaluating all isolates using a Qubit 2.0 fluorometer (Life Technologies Corp, Carlsbad, CA) (313).

Illumina Library Construction and Sequencing

Small insert PE libraries were constructed using the TruSeq Nano LT Library Prep Kit (Illumina #FC-121-4001) according to the standard protocol supplied by the manufacturer. All quail PE libraries were multiplexed and processed using PE-125 cycle runs (2×125 bp), with data generation (i.e., image processing and base calling) occurring in real time on the Illumina HiSeq 2500v4 High Output instrument (Illumina Inc., San Diego, CA). Briefly, the sequencing strategy consisted of multiplexing four barcoded birds per lane, which delivered high mitochondrial coverage across all individual quail.

Bobwhite Mitogenome Reference Mapping and Variant Detection

Prior to reference mapping, all Illumina sequence reads were trimmed for quality and adapter sequences using the CLC Genomics Workbench, as previously described (313). Comparison of the initial bobwhite draft mitogenome reference sequence (GenBank Accession AWGT00000000.1) (313) with recently published mitogenome reference

sequences for several members of the family Odontophoridae (New World quail) (314) revealed an in-frame gap in the *ND5* reference sequence; a complication related to *de novo* assembly of a circular mitochondrial chromosome into a linear contig (i.e., end gaps). However, the recent increase in mitogenome taxon sampling for species of the Odontophoridae (314) allowed for comparative correction and read-based validation of the gap within the previously reported bobwhite reference mitogenome sequence (GenBank Accession AWGT00000000.1) (313). Thereafter, the trimmed Illumina sequence reads generated for 51 bobwhites were individually mapped to the corrected bobwhite mitogenome reference sequence (313) (i.e., equivalent in length to GenBank Accession KJ914548.1; mitogenome = 16, 702 bp) using the reference mapping algorithm within the CLC Genomics Workbench (v7.0.3 and v7.5.1) (313). Reference mapping parameters were as follows: no masking; mismatch cost = 2; insertion cost = 3; deletion cost = 3; minimum read length fraction = 0.95; minimum fraction of nucleotide identity (similarity) = 0.95. Duplicate mapped reads were removed using the CLC Genomics duplicate mapped reads removal algorithm (version 1.0; For PCR-based libraries), which also aims to collapse reads that are only distinguished by minority-branch sequencing errors (http://www.clcbio.com/files/usermanuals/Mapped_Duplicate_Reads_Removal_Plugin.pdf). Thereafter, we used the CLC probabilistic variant detection algorithm (315) to predict bobwhite mitogenome variants (SNVs, Indels, MNVs; http://www.clcbio.com/files/whitepapers/Variant_Caller_WP_web.pdf). This algorithm uses a Bayesian model and a Maximum Likelihood approach to calculate prior and error probabilities for the Bayesian model. These probabilities are used to determine the most

likely allele combination per nucleotide position, with a user specified probability threshold for variant prediction equal to 0.95 (313). Additional user specified settings for variant detection were as follows: ignore nonspecific matches = yes; ignore broken read pairs = no; minimum coverage = 4; variant probability ≥ 0.95; require variant in both forward and reverse reads = yes; maximum expected variants = 2; ignore quality scores = no. Resulting variant tracks for all 51 bobwhites were annotated by sequence overlap using mitogenome annotations previously described (313, 314), and the functional consequences of all putative variants were predicted using the vertebrate mitochondrial genetic code implemented in the CLC Genomics Workbench (v7.0.3 and v7.5.1). Consensus mitogenome sequences for all 51 bobwhites were manually constructed using individual variant reports, and then compared for accuracy to the consensus sequences computed by the CLC Genomics Workbench (v7.0.3 and v7.5.1).

Generation of a Complete Scaled Quail Mitogenome Sequence

Two basic approaches were used to produce a complete scaled quail mitogenome. First, we mapped Illumina PE reads generated for a female scaled quail onto a bobwhite mitogenome reference sequence (GenBank Accessions AWGT00000000.1 and KJ914548.1) (313, 314) using the reference mapping algorithm implemented within the CLC Genomics Workbench (v7.5.1), and subsequently computed a scaled quail consensus. For comparison, all of the mapped reads were extracted and used in conjunction with the CLC *de novo* assembly algorithm, as previously described (313). Both approaches produced identical results (16,701 bp; GenBank Accession KT722338), and the scaled

quail mitogenome was comparatively annotated by sequence overlap(s) with mitogenome annotations previously described (313, 314). Sequence overlaps were established by alignments performed in the CLC genomics workbench (v7.5.1) and/or via ClustalW online (http://www.ebi.ac.uk/Tools/msa/clustalw2/), with subsequent manual inspection. Translation of all 13 mitogenome protein coding genes provided no evidence for frameshifts, premature stop codons, or missing data (i.e., gaps).

Population, Demographic, and Phylogenetic Analyses

For population analyses, all bobwhite mitogenome sequences were aligned using a multiple sequence alignment algorithm implemented within the CLC Genomics Workbench (v7.5.1; CLC alignment algorithm). This alignment included one bobwhite mitogenome sequence of unknown geographic origin and/or putative subspecies designation (GenBank Accession KJ914548.1) (314) as well as the gap-corrected (*ND5*) reference mitogenome (GenBank Accession AWGT00000000.1) (313). For bobwhites displaying unequivocal evidence of two intra-individual mitogenome haplotypes that differed by only one nucleotide (i.e., one heteroplasmic SNV; 10 / 13 bobwhites), both representative haplotype sequences were included in the alignment. Likewise, for bobwhites that displayed evidence for more than one heteroplasmic SNV (n = 3 bobwhites), we used allele frequency data to deduce the two putative intra-individual mitogenome haplotype sequences. Specifically, in all but one instance, the reference allele was observed at higher frequencies (i.e., by read count and coverage) than the alternative allele, thereby supporting the presence of one reference haplotype, and one alternative

haplotype comprised of minor alleles. The only heteroplasmic SNV that deviated from this trend was excluded from further analyses. Likewise, two SNVs that could not be unambiguously excluded as potential numts were also removed from further analyses (313). Bobwhite median joining haplotype networks (341) were constructed using Network 4.6.1.3 (Fluxus Technology Ltd, Suffolk, England), with the default character weights for SNPs and indels (10 and 20, respectively). All networks were visualized and annotated within Network Publisher (Fluxus Technology Ltd, Suffolk, England), with manual adjustment of branch angles to ensure proper magnification and clarity without changing the branch lengths. Tests of population differentiation and subdivision (K_S, K_S^*, Z, Z^*) (342), estimates of haplotype and nucleotide diversity (Hd; π) (238), and bobwhite demographic models (i.e., constant stable population vs growth-decline) (357-359, 367) were computed in DnaSP version 5.10.01 (368). Frequency distribution tests (D; F_S) (360, 361) were also performed in DnaSP version 5.10.01 (368), with the significance of the test assessed by the beta distribution (D) and/or via coalescent simulation (D; F_S) with 16,000 replicates. All computations performed in DnaSP version 5.10.01 (368) were based on the total number of mutations (excluding gaps), with one exception; the population growth-decline model was evaluated via mismatch distribution (i.e., pairwise number of differences) and the site frequency spectrum (i.e., segregating sites) (357-359, 367). Pairwise fixation index (F_{ST}) values were computed to assess genetic differentiation using a distance matrix in conjunction with a Tamura and Nei (369) model within the program ARLEQUIN v3.5.1.2 (370) (i.e., partial and complete D-loop; complete mitogenome). Both multi-locus and standard MKT's were performed using the available web interface

(http://mkt.uab.es/mkt/help_mkt.asp) with Jukes-Cantor correction (362, 363), with standard MKT's (363) for individual mitochondrial protein coding genes also calculated in DnaSP version 5.10.01 (368). Tajima's relative rate test (356) was performed within the software program Mega v6.0 (371), and maximum likelihood phylogenies were constructed with RAxML 7.2.8 (366) using a GTR+Γ model of sequence evolution, with bootstrap support values based on 1,000 pseudoreplicates.

CHAPTER VI

CONCLUSIONS AND FUTURE RESEARCH*

The bobwhite has increasingly been considered a well-suited model wildlife species for climate change, land use, toxicology, and conservation studies (67, 73, 74, 84, 147-152) and has been one of the most intensively researched and managed wildlife species in North America (67, 84, 147). However, like many other important wildlife species, the bobwhite has historically suffered from a paucity of genome-wide sequence and polymorphism data, which limited the implementation of genomic approaches for addressing key biological questions (153, 154). The research, results and deliverable reported in this dissertation provides analyses that will inform managers and conservation biologists about genetic or genomic support for existing subspecies designations, gene flow, population structure, and aspects of historic effective population size.

The ability to rapidly generate low-cost, high quality avian draft *de novo* genome assemblies in conjunction with coalescent models to reconstruct the demographic histories of species which are currently in decline provides a modern framework for understanding and monitoring both historic and recent population trends. Although the bobwhite has clearly declined across much of its native range (73, 74, 77-80), our estimates of N_e up until about 9-10 kya demonstrate that genomic diversity has remained quite high despite a substantial, historic bottleneck (Fig. 4). The same cannot be said for the scarlet macaw

* Portions of this chapter were reprinted with permission from "A Draft *De Novo* Genome Assembly for the Northern Bobwhite (*Colinus virginianus*) Reveals Evidence for a Rapid Decline in Effective Population Size Beginning in the Late Pleistocene" by Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, Rollins D, Tizard IR, Brightsmith DJ, Peterson MJ, Taylor JF, Seabury CM, 2014. PLoS ONE 9(3): e90240. doi:10.1371/journal.pone.0090240, 2014 Halley et al.

(Fig. 4), with our analyses indicating that N_e for the scarlet macaw was never as large as the bobwhite (Fig. 4), with the large disparity in effective population sizes between these two highly divergent species most likely a product of their opposing natural selection strategies (i. e., *r*- versus *K*-selection). Short generation times and large clutches in the bobwhite provide more opportunities for the creation of genomic diversity via meiotic recombination and new mutation than do the long generation times, small clutches, and very small broods for the scarlet macaw (2, 57, 250-252, 281). Therefore, our observations are concordant with genomic signatures of selection created by how opposing selection strategies (i.e., skewness in the *r*- versus *K*-selection continuum) would be expected to shape genomic diversity and the corresponding effective population sizes in these species (57, 251).

Considering the conclusions of human GWAS studies (i.e., genes, noncoding regions underlying quantitative trait loci), the results of our whole-genome analyses of divergence were often consistent with several fundamental biological differences noted between three divergent avian species, with independent replication of some outlier loci and trait classes that were previously suggested to be important among avian species (154). We also identified several potential candidate genes and noncoding regions which coincide with human GWAS studies for biological traits that appear disparate among the three investigated bird species, but also found previously reported evidence for purifying selection operating on some of the same genes we identified within our conserved outlier contigs (Table S11) [see Appendix A]. As described for a recent analysis of the scarlet macaw genome, the overwhelming majority of the bobwhite contigs (NB1.0) classified as

outliers for divergence with the chicken and zebra finch were determined to contain noncoding sequences, which is consistent with the hypothesis that noncoding regions of the genome are likely to underlie differences in species-specific genome regulation and traits (154, 192-195).

For bobwhite samples included in the present study, utilization of small, popular mitochondrial fragments (9, 336-340) were observed to be largely insufficient to elucidate the true mitogenome haplotype structure and corresponding levels of divergence within a structured population. Likewise, discordant demographic inferences and failure to detect the extent of population substructure were also possible, as evidenced by small versus complete mitogenome sequence analyses. Our analyses of complete mitogenome sequence data for 53 bobwhites from six ecoregions across two U.S. states supported the potential for perhaps two putative subspecies that did not adhere to prior geographic designations (5, 9, 10), with the molecular caveat being that future nuclear genome analyses are also necessary to fully clarify bobwhite population structure and putative subspecies designations west of the Mississippi River (USA). Collectively, our analyses of bobwhite mitogenomic data, including evidence for heteroplasmy, strongly support the deployment of low-cost, high-yielding, next-generation sequencing technologies in place of conventional PCR-based analyses of small mitochondrial fragments for future population studies. Additionally, population, demographic, and phylogenetic analyses reported in this study were robust to the inclusion or exclusion of heteroplasmic SNVs [see Appendix C-D].

To date, few genetic studies that focus on genetic variability within and among bobwhite populations have emerged, with none of these studies taking a whole genome approach (9, 21, 372). Herein, we used complete mitogenome information in our analyses, which brought new clarity to issues like genetic support for bobwhite geographic subspecies designation, population structure, and the historical demography of maternal lineages. Notably, bobwhites are naturally subdivided (i.e., to some degree) into social coveys that generally occupy a limited geographic area (2, 14, 21-23). However, as demonstrated by radio telemetry studies, bobwhites are also known to disperse (i.e., spring dispersal, breeding season, brood movements) (138, 373, 374), which is theorized to enable bobwhite range expansions, which are likely important during times of draught (375). The dispersal of species from fragmented habitats is also hypothesized to potentially help increase gene flow between subpopulations and reduce the likelihood of inbreeding (373, 374, 376, 377). However, as noted by our complete mitogenome analyses, little gene flow is apparent among the two bobwhite maternal lineages detected (Figures with FST, tables, etc.), which may indicate low dispersal due to extensive habitat fragmentation (i.e., lack of suitable nearby habitats for expansion), with significant population substructure related to suitable (i.e., pocketed) habitats.

At present a need remains to examine the nuclear genetic relationships and overall levels of divergence within and between putative bobwhite subspecies; to assess whether or not nuclear variation and divergence also supports the potential for two putative subspecies in our samples, and to evaluate the concordance of nuclear and mitogenome divergence. At present, we have generated in excess of 120 bobwhite nuclear genomes

from six discrete ecoregions across TX and OK. These data will be instrumental in providing fine scale resolution regarding population structure, and for evaluating whether or not natural selection has significantly altered the distribution of polymorphism within wild bobwhite genomes. Importantly, naturally occurring infectious diseases have always been considered a potential factor contributing to bobwhite decline, although very few reports provide overt evidence for this. Nevertheless, the nuclear genome itself will tell a story about the history of the bobwhite, and future studies aimed at evaluating the distribution (i.e., immune-related genes) and magnitude of natural selection may elucidate important aspects of bobwhite natural history and the current broad-scale declines. Finally, while this dissertation provides a basis for instituting both nuclear and mitogenomic data that could be used for addressing key biological and conservational questions in the bobwhite, the limited number of birds/populations sampled in our initial studies does provide a shortcoming. By expanding the sampling scope across more states, ecoregions, putative subspecies and subpopulations, additional information regarding population demography, gene flow, and genetic support for subspecies status will emerge on a much greater scale; which could positively augment current management practices via genomically informed translocation, and/or by monitoring genetic diversity among pen-reared breeding stocks.

LITERATURE CITED

1. Stokes AW. Behavior of the bobwhite, *Colinus virginianus*. Auk. 1967;84(1):1-33.

2. Brennan LA. Texas quails: ecology and management: Texas A&M University Press; 2007.

3. Judd SD. The bobwhite and other quails of the United States in their economic relations: US Department of Agriculture; 1905.

4. Shanaway M. Quail production systems: a review: Food & Agriculture Organization; 1994.

5. del Hoyo J. Handbook of the birds of the world. Vol. 2, New World vultures to guineafowl: Lynx edicions; 1994.

6. Brennan LA. Northern Bobwhite (*Colinus virginianus*). The Birds of North America. 1999(397):28.

7. Guthery FS. Beef, brush, and bobwhites: quail management in cattle country: Texas A&M University Press; 1986.

8. Ashton KG. Patterns of within-species body size variation of birds: strong evidence for Bergmann's rule. Global Ecology and Biogeography. 2002;11(6):505-23.

9. Williford D, Deyoung RW, Honeycutt RL, Brennan LA, Hernández F, Wehland EM, et al. Contemporary genetic structure of the northern bobwhite west of the Mississippi River. The Journal of Wildlife Management. 2014;78(5):914-29.

10. Madge S, McGowan P. Pheasants, Partridges, and Grouse: A Guide to the Pheasants, Partridges, Quails, Grouse, Guineafowl, Buttonquails, and Sandgrouse of the World. Princeton University Press, Princeton, New Jersey, USA. 2002.

11. Hernandez F, Kuvlesky Jr WP, DeYoung RW, Brennan LA, Gall SA. Recovery of rare species: case study of the masked bobwhite. Journal of Wildlife Management. 2006;70(3):617-31.

12. Bailey ED, Baker JA. Recognition Characteristics in Covey Dialects of Bobwhite Quail. The Condor. 1982;84(3):317-20.

13. Kuvlesky W, Swank W, Silvy N, editors. Habitat selection of northern bobwhite in the Rio Grande plains of Texas. Quail V: proceedings of the fifth national quail symposium (SJ DeMaso, WP Kuvlesky, F Hernández, and M E Berger, editors) Texas Parks and Wildlife Department, Austin; 2002.

14. Faircloth BC. An integrative study of social and reproductive systems in northern bobwhite (*Colinus virginianus*): a non-migratory, avian species bearing precocial young: University of Georgia; 2008.

15. Stoddard HL. The bobwhite quail: its habits, preservation and increase: Charles Scribner's Sons, New York, New York; 1931.

16. Rosene W. The bobwhite quail, its life and management: Rutgers University Press, New Brunsqick, New Jersey; 1969.

17. Lehmann VW. Bobwhites in the Rio Grande plain of Texas: Texas A & M University Press; 1984.

18. Wilkins RN. Influence of grazing management on population attributes, habitats, and habitat selection of bobwhites in south Texas: Texas A&M University; 1987.

19. Carver AV, Burger Jr LW, Palmer WE, Brennan LA, editors. Vegetation characteristics in seasonal-disked fields and at bobwhite brood locations. Proceedings of the Annual Conference Southeastern Association Fish and Wild Agencies, Louisville, KY, USA; 2001.

20. Hiller TL, Guthery FS. Microclimate versus predation risk in roost and covert selection by bobwhites. Journal of Wildlife Management. 2005;69(1):140-9.

21. Ellsworth DL, Roseberry JL, Klimstra W. Genetic structure and gene flow in the Northern Bobwhite. The Auk. 1989:492-5.

22. Murphy A, Baskett TS. Bobwhite mobility in central Missouri. The Journal of Wildlife Management. 1952:498-510.

23. Lewis JB. Further studies of bob-white mobility in central Missouri. The Journal of Wildlife Management. 1954:414-6.

24. Case RM. Bioenergetics of a covey of bobwhites. The Wilson Bulletin. 1973:52-9.

25. Pulliam HR, Millikan GC. Social organization in the nonreproductive season. Avian biology. 1982;6:169-97.

26. Williams CK, Lutz RS, Applegate RD. Optimal group size and northern bobwhite coveys. Animal Behaviour. 2003;66(2):377-87.

27. Taylor W. The Bobwhite. Austin, Texas: The Univeristy Extrension Campus; 1917.

28. Errington PL. Vulnerability of bob-white populations to predation. Ecology. 1934;15(2):110-27.

29. Guthery FS, Hiller TL, Puckett Jr WH, Baker RA, Smith SG, Rybak AR. Effects of feeders on dispersion and mortality of bobwhites. Wildlife Society Bulletin. 2004;32(4):1248-54.

30. Roseberry JL, Klimstra WD. Population ecology of the bobwhite: Southern Illinois University Press; 1984.

31. Taylor JS, Church KE, Rusch DH, Cary JR. Macrohabitat effects on summer survival, movements, and clutch success of northern bobwhite in Kansas. The Journal of wildlife management. 1999:675-85.

32. Cook MP. Northern bobwhite breeding season dispersal, habitat use, and survival in a southeastern agricultural landscape: Thesis, University of Georgia, Athens, USA; 2004.

33. Terhune TM, Sisson DC, Stribling HL. The efficacy of relocating wild northern bobwhites prior to breeding season. Journal of Wildlife Management. 2006;70(4):914-21.

34. Burger Jr LW, Dailey TV, Kurzejeski EW, Ryan MR. Survival and cause-specific mortality of northern bobwhite in Missouri. The Journal of wildlife management. 1995:401-10.

35. Simpson R, editor A study of bobwhite quail nest initiation dates, clutch sizes, and hatch sizes in southwest Georgia. Proceedings of the First National Bobwhite Quail Symposium, Oklahoma State University Stillwater, Oklahoma; 1972.

36. Klimstra W. Bob-white quail nesting and production in southeastern Iowa. Iowa State College Journal of Science. 1950;24(4):385-95.

37. Klimstra WD, Roseberry JL. Nesting ecology of the bobwhite in southern Illinois. Wildlife Monographs. 1975:3-37.

38. Rosenberry JL. Avian Mortality in Southern Illinois Restulting from Severe Weather Conditions. Ecology. 1962;43(4):739-40.

39. Burger Jr LW, Ryan MR, Dailey TV, Kurzejeski EW. Reproductive strategies, success, and mating systems of northern bobwhite in Missouri. The Journal of wildlife management. 1995:417-26.

40. Curtis P, Mueller B, Doerr P, Robinette C, editors. Seasonal survival of radio-marked northern bobwhite quail from hunted and non-hunted populations. Proceedings of the International Biotelemetry Symposium; 1988.

41. Curtis P, Mueller B, Doerr P, Robinette C, DeVos T, editors. Potential polygamous breeding behavior in northern bobwhite. Proceedings of the National Quail Symposium; 1993.

42. Rolland V, Hostetler JA, Hines TC, Percival HF, Oli MK. Factors influencing reproductive performance of northern bobwhite in South Florida. European Journal of Wildlife Research. 2011;57(4):717-27.

43. Emlen ST, Oring LW. Ecology, sexual selection, and the evolution of mating systems. Science. 1977;197(4300):215-23.

44. Hernandez F, Peterson MJ. Northern bobwhite ecology and life history. Texas Quails Texas A&M Press, College Station. 2007:40-64.

45. Guthery FS, Koerth NE, Smith DS. Reproduction of northern bobwhites in semiarid environments. The Journal of Wildlife Management. 1988:144-9.

46. Staller EL, Palmer WE, Carroll JP, Thornton RP, Sisson DC. Identifying predators at northern bobwhite nests. Journal of Wildlife Management. 2005;69(1):124-32.

47. DeVos T, Mueller B, editors. Reproductive ecology of northern bobwhite in north Florida. Proceedings of the National Quail Symposium; 1993.

48. Puckett KM, Palmer WE, Bromley PT, Anderson Jr JR, Sharpe TL, editors. Bobwhite nesting ecology and modern agriculture: a management experiment. Proceedings of the Annual Conference of the Southeastern Association of Fish and Wildlife Agencies; 1995.

49. Rollins D, Carroll JP. Impacts of predation on northern bobwhite and scaled quail. Wildlife Society Bulletin. 2001;29(1):39-51.

50. Fies M, Puckett K, editors. Depredation patterns of northern bobwhite nest predators in Virginia. Proceedings of the National Quail Symposium; 2000.

51. Rader MJ, Teinert TW, Brennan LA, Hernandez F, Silvy NJ, WU X. Identifying predators and nest fates of bobwhites in southern Texas. The Journal of wildlife management. 2007;71(5):1626-30.

52. Suchy W, Munkel R, editors. Breeding strategies of the northern bobwhite in marginal habitat. Proceedings of the National Quail Symposium; 1993.

53. Farnsworth GL, Simons TR, Brawn J. How many baskets? Clutch sizes that maximize annual fecundity of multiple-brooded birds. The Auk. 2001;118(4):973-82.

54. Baicich P, C.J. 0. HARRISON. A guide to the nests, eggs, and nestlings of North American birds. Academic Press, San Diego, California. 1997

55. Sermons WO, Speake DW. Production of second broods by northern bobwhites. The Wilson bulletin. 1987;99(2):285-6.

56. Brazil KA, Brennan LA, Hernàndez F, Ballard BM, Bryant FC. Order and Chaos: Northern bobwhite porductivity and nest-habitat relationships in south Texas. Texas Ornithological Society. 1970:592-7.

58. Starck JM, Ricklefs RE. Avian growth and development: evolution within the altricial-precocial spectrum: Oxford University Press; 1998.

59. Eubanks TR, Dimmick RW. Dietary patterns of bobwhite quail on Ames Plantation. University of Tennessee Agricultural Experiment Station Bulletin; 1974.

60. Choi I-H, Ricklefs RE, Shea RE. Skeletal muscle growth, enzyme activities, and the development of thermogenesis: a comparison between altricial and precocial birds. Physiological Zoology. 1993:455-73.

61. Lochmiller RL, Vestey MR, Boren JC. Relationship between protein nutritional status and immunocompetence in northern bobwhite chicks. The Auk. 1993:503-10.

62. Hurst GA, editor Insects and bobwhite quail brood habitat management. Proceedings of the National Quail Symposium; 1972.

63. Nestler R, Bailey W, McClure HE. Protein requirements of bobwhite quail chicks for survival, growth, and efficiency of feed utilization. The Journal of Wildlife Management. 1942:185-93.

64. Williams HW, Stokes A, Wallen JC. The food call and display of the bobwhite quail (*Colinus virginianus*). The Auk. 1968:464-76.

65. Nice MM. Food of the bobwhite. Journal of Economic Entomology. 1910;3(3):295-313.

66. Nestler RB. Nutrition of bobwhite quail. The Journal of Wildlife Management. 1949:342-58.

67. Lusk JJ, Guthery FS, George RR, Peterson MJ, DeMaso SJ. Relative abundance of bobwhites in relation to weather and land use. The Journal of wildlife management. 2002:1040-51.

68. Bridges AS, Peterson MJ, Silvy NJ, Smeins FE, Ben Wu X. Differential influence of weather on regional quail abundance in Texas. The Journal of wildlife management. 2001:10-8.

69. Hernández F, Hernández F, Arredondo JA, Bryant FC, Brennan LA, Bingham RL. Influence of precipitation on demographics of northern bobwhites in southern Texas. Wildlife Society Bulletin. 2005;33(3):1071-9.

70. Leopold A. Report on a game survey of the north central states. Sporting Arms and Ammunition Manufactures Institute, Madison, Wis 1931.

71. Errington PL, Hamerstrom F. northern bob-white's winter territory. Iowa Agricultural Experiment Station Research Bulletin. 1936; 201:301-443.

72. Lehmann VW. Increase quail by improving their habitat. Texas Game, Fish and Oyster Commission, Austin. 1937.

73. Brennan LA. How can we reverse the northern bobwhite population decline? Wildlife Society Bulletin. 1991:544-55.

74. Sauer JR, Link WA, Nichols JD, Royle JA. Using the North American Breeding Bird Survey as a tool for conservation: a critique of Bart et al.(2004). Journal of Wildlife Management. 2005;69(4):1321-6.

75. Droege S, Sauer JR, editors. Northern bobwhite, gray partridge, and ring-necked pheasant population trends (1966–1988) from the North American Breeding Bird Survey. Perdix V: Gray partridge and ring-necked pheasant workshop; 1990: Kansas Department of Wildlife and Parks Emporia, Kansas.

76. Church K, Sauer J, Droege S, editors. Population trends of quails in North America. Proceedings of the National Quail Symposium; 1993.

77. Brady S, Flather C, Church K. Range-wide declines of northern bobwhite (*Colinus virginianus*): land use patterns and population trends. Gibier Faune Sauvage. 1998;15:413-31.

78. Peterson M, Wu X, Rho P, editors. Rangewide trends in landuse and northern bobwhite abundance: an exploratory analysis. Proceedings of the National Quail Symposium; 2002.

79. Sauer J, Hines J, Fallon J, Pardieck K, Ziolkowski DJ, al. e. The North American breeding bird survey. Results and analysis. 2012;2003.

80. Hernández F, Brennan LA, DeMaso SJ, Sands JP, Wester DB. On reversing the northern bobwhite population decline: 20 years later. Wildlife Society Bulletin. 2013;37(1):177-88.

81. Terhune TM, Sisson DC, Stribling HL, Carroll JP. Home range, movement, and site fidelity of translocated northern bobwhite (*Colinus virginianus*) in southwest Georgia, USA. European Journal of Wildlife Research. 2006;52(2):119-24.

82. Guthery F, Forrester N, Nolte K, Cohen W, Kuvlesky Jr W, editors. Potential effects of global warming on quail populations. Proceedings of the National Quail Symposium; 2000.

83. Reyna K, Burggren W. Upper lethal temperatures of Northern Bobwhite embryos and the thermal properties of their eggs. Poultry science. 2012;91(1):41-6.

84. Williams CK, Guthery FS, Applegate RD, Peterson MJ. The northern bobwhite decline: scaling our management for the twenty-first century. Wildlife Society Bulletin. 2004;32(3):861-9.

85. Mueller JM, Dabbert CB, Demarais S, Forbes AR. Northern bobwhite chick mortality caused by red imported fire ants. The Journal of wildlife management. 1999:1291-8.

86. Allen CR, Willey R, Myers P, Horton P, Buffa J. Impact of red imported fire ant infestation on northern bobwhite quail abundance trends in southeastern United States. Journal of Agricultural and Urban Entomology. 2000;17:43-51.

87. Myers PE, Allen CR, Birge HE. Consuming Fire Ants Reduces Northern Bobwhite Survival and Weight Gain 1. Journal of Agricultural and Urban Entomology. 2014;30(1):49-58.

88. Ottinger M, Quinn M, Lavoie E, Abdelnabi M, Thompson N, Hazelton J, et al. Consequences of endocrine disrupting chemicals on reproductive endocrine function in birds: establishing reliable end points of exposure. Domestic animal endocrinology. 2005;29(2):411-9.

89. Kitulagodage M, Isanhart J, Buttemer WA, Hooper MJ, Astheimer LB. Fipronil toxicity in northern bobwhite quail *Colinus virginianus*: reduced feeding behaviour and sulfone metabolite formation. Chemosphere. 2011;83(4):524-30.

90. Peterson M, Perez R, editors. Is quail hunting self regulatory? Northern bobwhite and scaled quail abundance and quail hunting in Texas. Proceedings of the National Quail Symposium; 2000.

91. Peterson MJ. Northern bobwhite and scaled quail abundance and hunting regulation: a Texas example. The Journal of wildlife management. 2001:828-37.

92. Williams CK, Lutz RS, Applegate RD, Palmer W. Winter survival and additive harvest in northern bobwhite coveys in Kansas. Journal of Wildlife Management. 2004;68(1):94-100.

93. Xiang L, Guo F, Zhang H, LaCoste L, Rollins D, Bruno A, et al. Gene discovery, evolutionary affinity and molecular detection of *Oxyspirura petrowi*, an eye worm parasite of game birds. BMC microbiology. 2013;13(1):233.

94. Dunham NR, Soliz LA, Fedynich AM, Rollins D, Kendall RJ. Evidence of an Oxyspirura petrowi epizootic in Northern Bobwhites (*Colinus virginianus*), Texas, USA. Journal of wildlife diseases. 2014;50(3):552-8.

95. Bruno A, Fedynich AM, Smith-Herron A, Rollins D. Patholigical response of northern bobwhites to *oxspirura pterowi* infections. Journal of Parasitology. 2015.

96. Dunham NR, Kendall RJ. Evidence of *Oxyspirura petrowi* in Migratory Songbirds Found in the Rolling Plains of West Texas, USA. Journal of wildlife diseases. 2014;50(3):711-2.

97. DeVos Jr T, Speake DW. Effects of releasing pen-raised northern bobwhites on survival rates of wild populations of northern bobwhites. Wildlife Society Bulletin. 1995:267-73.

98. Terhune TM, Sisson DC, Palmer WE, Faircloth BC, Stribling HL, Carroll JP. Translocation to a fragmented landscape: survival, movement, and site fidelity of Northern Bobwhites. Ecological Applications. 2010;20(4):1040-52.

99. Scott JL, Hernández F, Brennan LA, Ballard BM, Janis M, Forrester ND. Population demographics of translocated northern bobwhites on fragmented habitat. Wildlife Society Bulletin. 2013;37(1):168-76.

100. Baumgartner FM. Dispersal and survival of game farm bobwhite quail in northcentral Oklahoma. The Journal of Wildlife Management. 1944:112-8.

101. Buechner HK. An evaluation of restocking with pen-reared bobwhite. The Journal of Wildlife Management. 1950:363-77.

102. Evans K, Smith M, Burger Jr L, Chambers R, Houston A, Carlisle R. Release of pen-reared bobwhites: potential consequences to the genetic integrity of resident wild populations. Gamebird Georgia: University of Georgia. 2006:121-33.

103. Mayer AM. Bob White: the game bird of America. Whitefish, Montana: Kessinger Pub Co; 1883.

104. Burger LW, Miller DA, Southwick RI. Economic impact of northern bobwhite hunting in the southeastern United States. Wildlife Society Bulletin. 1999:1010-8.

105. Hernández F, Guthery FS. Beef, brush, and bobwhites: quail management in cattle country: Texas A&M University Press; 2012.

106. Kinsey JC. Effectiveness of Surrogators as a propagation tool for northern bobwhites in south-central Texas: Texas State University-San Marcos; 2011.

107. Sands JP, Schnupp MJ, Teinert TW, DeMaso SJ, Hernandez F, Brennan LA, et al. Tests of an additive harvest mortality model for northern bobwhite *Colinus virginianus* harvest management in Texas, USA. Wildlife Biology. 2013;19(1):12-8.

108. Dodd EP, Bryant FC, Brennan LA, Gilliland C, Dudensing R, McCorkle D. An Economic Impact Analysis of South Texas Landowner Hunting Operation Expenses. Journal of Fish and Wildlife Management. 2013;4(2):342-50.

109. Brennan L, editor A decade of progress, a decade of frustration. National Quail Symposium Proceedings; 2002.

110. Leopold A. Report on a game survey of Mississippi: Mississippi Department of Wildlife, Fisheries and Parks, Jackson; 1929.

111. Sandercock BK, Jensen WE, Williams CK, Applegate RD. Demographic sensitivity of population change in northern bobwhite. Journal of Wildlife Management. 2008;72(4):970-82.

112. Leopold A. Game management: University of Wisconsin Press; 1987.

113. Errington PL, Hamerstrom Jr F. Bob-white winter survival on experimentally shot and unshot areas. Iowa State College Journal of Science. 1935;9:625-39.

114. Errington PL. Some contributions of a fifteen-year local study of the northern bobwhite to a knowledge of population phenomena. Ecological Monographs. 1945:2-34.

115. Errington PL. Mobility of the northern bob-white as indicated by banding returns. Bird-Banding. 1933;4(1):1-7.

116. Pollock KH, Moore CT, Davidson WR, Kellogg FE, Doster GL. Survival rates of bobwhite quail based on band recovery analyses. The Journal of Wildlife Management. 1989:1-6.

117. Guthery FS. A philosophy of habitat management for northern bobwhites. The Journal of wildlife management. 1997:291-301.

118. Scott TG. Bobwhite thesaurus. International quail foundation, Edgefield, South Carolina; 1985.

119. Stockard CR. Nesting habits of birds in Mississippi. The Auk. 1905;22(2):146-58.

120. Bangs O. Notes on a collection of Bahama birds. The Auk. 1900;17(3):283-93.

121. Jones L. Winter Birds. The Wilson Bulletin. 1902:129-32.

122. Kopman HH. List of Birds Seen in Jefferson Parish, Louisiana, April 1, 1904. The Auk. 1905:140-5.

123. Howell AH. Notes on the summer birds of Kentucky and Tennessee. The Auk. 1910:295-304.

124. Lack D. Courtship feeding in birds. The Auk. 1940:169-78.

125. Lack D. Pair-formation in birds. The Condor. 1940:269-86.

126. Leopold AS. Age determination in quail. The Journal of Wildlife Management. 1939:261-5.

127. Hernández F, Guthery FS, Kuvlesky Jr WP. The legacy of bobwhite research in south Texas. The Journal of wildlife management. 2002:1-18.

128. Lehmann VW. Mobility of bobwhite quail in southwestern Texas. The Journal of Wildlife Management. 1946:124-36.

129. Lehmann VW. Bobwhite quail reproduction in southwestern Texas. The Journal of Wildlife Management. 1946:111-23.

130. Hanson HC. Criteria of age of incubated mallard, wood duck, and bob-white quail eggs. The Auk. 1954:267-72.

131. Caughley G. Analysis of vertebrate populations. Wiley, New York, New York; 1977.

132. Rosene W. A summer whistling cock count of bobwhite quail as an index to wintering populations. The Journal of Wildlife Management. 1957:153-8.

133. Bennitt R. Some aspects of Missouri quail and quail hunting, 1938-1948. Missouri Conservation Commission. Jefferson City, Missouri; 1951.

134. Norton HW, Scott TG, Hanson WR, Klimstra WD. Whistling-cock indices and bobwhite populations in autumn. The Journal of Wildlife Management. 1961:398-403.

135. DeMaso SJ, Guthery FS, Spears GS, Rice SM. Morning covey calls as an index of northern bobwhite density. Wildlife Society Bulletin. 1992:94-101.

136. Davis B. Effects of brush control on quail populations. Austin, Texas: Texas Parks and Wildlife; 1979.

137. Cantu R, Everett D, editors. Reproductive success and brood survival of bobwhite quail as affected by grazing practices. Proceedings of the National Quail Symposium; 1982.

138. Taylor JS, Guthery FS. Daily movements of Northern Bobwhite broods in southern Texas. The Wilson Bulletin. 1994:148-50.

139. Guthery FS. Bobwhites and brush control. Rangelands Archives. 1980;2(5):202-4.

140. Guthery F. Bobwhite and turkey management during drought. Livestock and Wildlife Management During Drought, RD Brown, ed Caesar Kleberg Wildlife Research Institute, Kingsville, Texas. 1984;1986:31-6.

141. Johnson DB, Guthery FS. Loafing coverts used by northern bobwhites in subtropical environments. The Journal of Wildlife Management. 1988:464-9.

142. Johnson DB, Guthery FS, Kane AH. Attributes of whistling posts used by northern bobwhites (*Colinus virginianus*). The Southwestern Naturalist. 1990:229-31.

143. Kopp SD, Guthery FS, Forrester ND, Cohen WE. Habitat selection modeling for northern bobwhites on subtropical rangeland. The Journal of wildlife management. 1998:884-95.

144. Beasom SL, editor Intensive short-term predator removal as a game management tool. Transactions of the North American Wildlife and Natural Resources Conference; 1974.

145. Guthery FS, Beasom SL. Responses of game and nongame wildlife to predator control in south Texas. Journal of Range Management. 1977:404-9.

146. Rice SM, Guthery FS, Spears GS, DeMaso SJ, Koerth BH. A precipitation-habitat model for northern bobwhites on semiarid rangeland. The Journal of wildlife management. 1993:92-102.

147. Quinn MJ, Hanna TL, Shiflett AA, McFarland CA, Cook ME, Johnson MS, et al. Interspecific effects of 4A-DNT (4-amino-2, 6-dinitrotoluene) and RDX (1, 3, 5-trinitro-1, 3, 5-triazine) in Japanese quail, Northern bobwhite, and Zebra finch. Ecotoxicology. 2013;22(2):231-9.

148. Johnson MS, Michie MW, Bazar MA, Gogal RM. Influence of oral 2, 4-dinitrotoluene exposure to the Northern Bobwhite (*Colinus virginianus*). International journal of toxicology. 2005;24(4):265-74.

149. Quinn MJ, Bazar MA, McFarland CA, Perkins EJ, Gust KA, Gogal RM, et al. Effects of subchronic exposure to 2, 6-dinitrotoluene in the northern bobwhite (*Colinus virginianus*). Environmental Toxicology and Chemistry. 2007;26(10):2202-7.

150. Quinn Jr MJ, McFarland CA, LaFiandra EM, Johnson MS. A preliminary assessment of relative sensitivities to foreign red blood cell challenges in the

northern bobwhite for potential evaluation of immunotoxicity. Journal of immunotoxicology. 2009;6(3):171-3.

151. Brausch JM, Blackwell BR, Beall BN, Caudillo C, Kolli V, Godard-Codding C, et al. Effects of polycyclic aromatic hydrocarbons in northern bobwhite quail (*Colinus virginianus*). Journal of Toxicology and Environmental Health, Part A. 2010;73(8):540-51.

152. Rawat A, Gust KA, Deng Y, Garcia-Reyero N, Quinn MJ, Johnson MS, et al. From raw materials to validated system: The construction of a genomic library and microarray to interpret systemic perturbations in Northern bobwhite. Physiological genomics. 2010;42(2):219-35.

153. Seabury CM, Bhattarai EK, Taylor JF, Viswanathan GG, Cooper SM, Davis DS, et al. Genome-wide polymorphism and comparative analyses in the white-tailed deer (*Odocoileus virginianus*): a model for conservation genomics. PLoS ONE. 2011;6(1):e15811.

154. Seabury CM, Dowd SE, Seabury PM, Raudsepp T, Brightsmith DJ, Liboriussen P, et al. A multi-platform draft de novo genome assembly and comparative analysis for the Scarlet Macaw (*Ara macao*). PLoS ONE. 2013;8(5):e62415.

155. Bec¸ak M, Bec¸ak W, Roberts F, Shoffner R, Volpe E. Aves. In: Benirschke K, Hsu TC, editors. Chromosome Atlas: Fish, Amphibians, Reptiles, and Birds: Springer-Verlag; 1971.

156. Hale DW, Ryder EJ, Sudman PD, Greenbaum IF. Application of synaptonemal complex techniques for determination of diploid number and chromosomal morphology in birds. The Auk. 1988:776-9.

157. Rawat A, Gust KA, Elasri MO, Perkins EJ. Quail Genomics: a knowledgebase for Northern bobwhite. BMC bioinformatics. 2010;11(Suppl 6):S13.

158. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences. 1977;74(12):5463-7.

159. Henson J, Tischler G, Ning Z. Next-generation sequencing and large genome assemblies. Pharmacogenomics. 2012;13(8):901-15.

160. Pop M. Genome assembly reborn: recent computational challenges. Briefings in bioinformatics. 2009;10(4):354-66.

161. Oleksyk TK, Pombert J-F, Siu D, Mazo-Vargas A, Ramos B, Guiblet W, et al. A locally funded Puerto Rican parrot (*Amazona vittala*) genome sequencing project increases avian data and advances young researcher education. Gigascience. 2012;1.

162. Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of species divergence in Ficedula flycatchers. Nature. 2012;491(7426):756-60.

163. Zhan X, Pan S, Wang J, Dixon A, He J, Muller MG, et al. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. Nature genetics. 2013;45(5):563-6.

164. Shapiro MD, Kronenberg Z, Li C, Domyan ET, Pan H, Campbell M, et al. Genomic diversity and evolution of the head crest in the rock pigeon. Science. 2013;339(6123):1063-7.

165. Wang B, Ekblom R, Bunikis I, Siitari H, Höglund J. Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. BMC genomics. 2014;15(1):180.

166. Miller AD, Good RT, Coleman RA, Lancaster ML, Weeks AR. Microsatellite loci and the complete mitochondrial DNA sequence characterized through next generation sequencing and de novo genome assembly for the critically endangered orange-bellied parrot, *Neophema chrysogaster*. Molecular biology reports. 2013;40(1):35-42.

167. Cai Q, Qian X, Lang Y, Luo Y, Xu J, Pan S, et al. Genome sequence of ground tit, *Pseudopodoces humilis*, and its adaptation to high altitude. Genome Biology. 2013;14(3):R29.

168. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 2004;432(7018):695-716.

169. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg LA, et al. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS biology. 2010;8(9):e1000475.

170. Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, et al. A Draft De Novo Genome Assembly for the Northern Bobwhite (*Colinus virginianus*) Reveals Evidence for a Rapid Decline in Effective Population Size

Beginning in the Late Pleistocene; 2014:(Supplemental tables). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0090240.

171. Tiersch T, Wachtel S. On the evolution of genome size of birds. Journal of Heredity. 1991;82(5):363-8.

172. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. Nature. 2013;497(7451):579-84.

173. Griffin DK, Robertson L, Tempest HG, Skinner BM. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenetic and genome research. 2007;117(1-4):64-77.

174. Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, et al. The dynamics of chromosome evolution in birds and mammals. Nature. 1999;402(6760):411-3.

175. Birds M. Above the Heads of Dinosaurs. University of California Press, Berkeley, CA. 2002.

176. Dyke G, Kaiser G. Living dinosaurs: the evolutionary history of modern birds: John Wiley & Sons, Hoboken, New Jersey; 2011.

177. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346(6215):1311-20.

178. Romanov MN, Tuttle EM, Houck ML, Modi WS, Chemnick LG, Korody ML, et al. The value of avian genomics to the conservation of wildlife. BMC genomics. 2009;10(Suppl 2):S10.

179. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006;22(23):2971-2.

180. Kumar S, Hedges SB. TimeTree2: species divergence times on the iPhone. Bioinformatics. 2011;27(14):2023-4.

181. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic acids research. 1999;27(23):4636-41.

182. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20(16):2878-9.

183. Dowd SE, Zaragoza J, Rodriguez JR, Oliver MJ, Payton PR. Windows. NET network distributed basic local alignment search toolkit (W. ND-BLAST). BMC bioinformatics. 2005;6(1):93.

184. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The genome of a songbird. Nature. 2010;464(7289):757-62.

185. Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, et al. The chicken B locus is a minimal essential major histocompatibility complex. Nature. 1999;401(6756):923-5.

186. Hughes CR, Miles S, Walbroehl JM. Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. Immunogenetics. 2008;60(5):219-31.

187. Balakrishnan CN, Ekblom R, Völker M, Westerdahl H, Godinez R, Kotkiewicz H, et al. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biology. 2010;8(1):29.

188. Ekblom R, Stapley J, Ball AD, Birkhead T, Burke T, Slate J. Genetic mapping of the major histocompatibility complex in the zebra finch (*Taeniopygia guttata*). Immunogenetics. 2011;63(8):523-30.

189. Monson M, Mendoza K, Velleman S, Strasburg G, Reed K. Expression profiles for genes in the turkey major histocompatibility complex B-locus. Poultry science. 2013;92(6):1523-34.

190. Casinos A, Cubo J. Avian long bones, flight and bipedalism. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology. 2001;131(1):159-67.

191. Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, et al. Genome-wide association study of body weight in chicken F2 resource population. PLoS ONE. 2011;6(7):e21872.

192. Meisler MH. Evolutionarily conserved noncoding DNA in the human genome: how much and what for? Genome research. 2001;11(10):1617-8.

193. Prabhakar S, Noonan JP, Pääbo S, Rubin EM. Accelerated evolution of conserved noncoding sequences in humans. Science. 2006;314(5800):786.

194. Pheasant M, Mattick JS. Raising the estimate of functional human sequences. Genome research. 2007;17(9):1245-53.

195. Johnson R, Samuel J, Ng CKL, Jauch R, Stanton LW, Wood IC. Evolution of the vertebrate gene regulatory network controlled by the transcriptional repressor REST. Molecular biology and evolution. 2009;26(7):1491-507.

196. Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, et al. Voxelwise genome-wide association study (vGWAS). Neuroimage. 2010;53(3):1160-74.

197. Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. PLoS ONE. 2009;4(8):e6501.

198. McClay JL, Adkins DE, Åberg K, Bukszár J, Khachane AN, Keefe RS, et al. Genome-wide pharmacogenomic study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. Neuropsychopharmacology. 2011;36(3):616-26.

199. Del Hoyo J, Elliott A, Christie D. Handbook of the Birds of the World. Vol. 15. Weavers to New World Warblers. British Birds. 2011;104:225-8.

200. Higgins PJ, Marchang S, Cowling S. Handbook of Australian, New Zealand and Antartic Birds, Boatbills to Starlings (Vol. 6). Melbourn; Oxford University Press. 2006:1132.

201. Mitgutsch C, Wimmer C, Sánchez-Villagra MR, Hahnloser R, Schneider RA. Timing of ossification in duck, quail, and zebra finch: intraspecific variation, heterochronies, and life history evolution. Zoological science. 2011;28(7):491.

202. Ringoen AR. Deposition of medullary bone in the female English sparrow, Passer domesticus (Linnaeus), and the Bobwhite quail, *Colinus virginianus*. Journal of morphology. 1945;77(2):265-83.

203. Dacke C, Arkle S, Cook D, Wormstone I, Jones S, Zaidi M, et al. Medullary bone and avian calcium regulation. The Journal of Experimental Biology. 1993;184(1):63-88.

204. Reynolds SJ. Uptake of ingested calcium during egg production in the zebra finch (*Taeniopygia guttata*). The Auk. 1997:562-9.

205. Blom J, Lilja C. A comparative study of embryonic development of some bird species with different patterns of postnatal growth. Zoology. 2005;108(2):81-95.

206. Murray JR, Varian-Ramos CW, Welch ZS, Saha MS. Embryological staging of the zebra finch, *Taeniopygia guttata*. Journal of morphology. 2013;274(10):1090-110.

207. Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nature genetics. 2010;42(12):1077-85.

208. Guthery F. On Bobwhites (Issue 27, W. L. Moody Jr. Natural Hisory Series). College Station: Texas A&M Univeristy Press. 2006:124.

209. Nager R, G L. The Zebra Finch. In: Hubrecht R, Kirkwood J, editors. The UFAW handbook on the care and management of laboratory and other research animals: John Wiley & Sons; 2010.

210. Mitchell GF, Verwoert GC, Tarasov KV, Isaacs A, Smith AV, Rietzschel ER, et al. Common Genetic Variation in the 3′-BCL11B Gene Desert Is Associated With Carotid-Femoral Pulse Wave Velocity and Excess Cardiovascular Disease Risk The AortaGen Consortium. Circulation: Cardiovascular Genetics. 2012;5(1):81-90.

211. Van Sligtenhorst I, Ding Z, Shi Z, Read R, Hansen G, Vogel P. Cardiomyopathy in α-Kinase 3 (*ALPK3*)–Deficient Mice. Veterinary Pathology Online. 2012;49(1):131-41.

212. Sotoodehnia N, Isaacs A, de Bakker PI, Dörr M, Newton-Cheh C, Nolte IM, et al. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. Nature genetics. 2010;42(12):1068-76.

213. Companioni O, Esparragón FR, Fernández-Aceituno AM, Pérez JCR. Genetic variants, cardiovascular risk and genome-wide association studies. Revista Española de Cardiología (English Edition). 2011;64(6):509-14.

214. Middelberg RP, Ferreira MA, Henders AK, Heath AC, Madden PA, Montgomery GW, et al. Genetic variants in *LPL*, *OASL* and *TOMM40*/*APOE*-*C1*-*C2*-*C4* genes are associated with multiple cardiovascular-related traits. BMC medical genetics. 2011;12(1):123.

215. Pfeufer A, Sanna S, Arking DE, Müller M, Gateva V, Fuchsberger C, et al. Common variants at ten loci modulate the QT interval duration in the QTSCD Study. Nature genetics. 2009;41(4):407-14.

216. Hagg S, Skogsberg J, Lundstrom J, Noori P, Nilsson R, Zhong H, et al. Multi-organ expression profiling uncovers a gene module in coronary artery disease involving transendothelial migration of leukocytes and LIM domain binding 2: the Stockholm

Atherosclerosis Gene Expression (STAGE) study. PLoS Genetics. 2009;5(12):e1000754.

217. Menzaghi C, Paroni G, De Bonis C, Coco A, Vigna C, Miscio G, et al. The protein tyrosine phosphatase receptor type f (*PTPRF*) locus is associated with coronary artery disease in type 2 diabetes. Journal of internal medicine. 2008;263(6):653-4.

218. Nolan DK, Sutton B, Haynes C, Johnson J, Sebek J, Dowdy E, et al. Fine mapping of a linkage peak with integration of lipid traits identifies novel coronary artery disease genes on chromosome 5. BMC genetics. 2012;13(1):12.

219. Newton-Cheh C, Eijgelsheim M, Rice KM, de Bakker PI, Yin X, Estrada K, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nature genetics. 2009;41(4):399-406.

220. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. Nature genetics. 2011;43(10):1005-11.

221. Artigas MS, Loth DW, Wain LV, Gharib SA, Obeidat Me, Tang W, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nature genetics. 2011;43(11):1082-90.

222. Hancock DB, Artigas MS, Gharib SA, Henry A, Manichaikul A, Ramasamy A, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. Plos Genetics. 2012;8(12), e1003098.

223. Egan MF, Straub RE, Goldberg TE, Yakub I, Callicott JH, Hariri AR, et al. Variation in *GRM3* affects cognition, prefrontal glutamate, and risk for schizophrenia. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(34):12604-9.

224. Kramer PL, Xu H, Woltjer RL, Westaway SK, Clark D, Erten-Lyons D, et al. Alzheimer disease pathology in cognitively healthy elderly: a genome-wide study. Neurobiology of aging. 2011;32(12):2113-22.

225. Ersland KM, Christoforou A, Stansberg C, Espeseth T, Mattheisen M, Mattingsdal M, et al. Gene-based analysis of regionally enriched cortical genes in GWAS data sets of cognitive traits and psychiatric disorders. PLoS ONE. 2012;7(2):e31687.

226. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang S-J, et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. BMC medical genetics. 2007;8(Suppl 1):S3.

227. Shetty PB, Hua T, Bamidele T, Morrison AC, Hanis CL, Rao DC, et al. Variants in *CXADR* and *F2RL1* are associated with blood pressure and obesity in African-Americans in regions identified through admixture mapping. Journal of hypertension. 2012;30(10):1970.

228. Eijgelsheim M, Newton-Cheh C, Sotoodehnia N, de Bakker PI, Müller M, Morrison AC, et al. Genome-wide association analysis identifies multiple loci related to resting heart rate. Human molecular genetics. 2010;19(19):3885-94.

229. Kung AW XS, Cherny S, Li GH, Gao Y, et al. Association of Stochastic Population Dynamics in Ecology and *JAG1* with bone mineral density and osteoporotic fractures: a genome-wide association study and followup replication studies. American Journal of Human Genetics. 2010;86:229.

230. Deng F-Y, Zhao L-J, Pei Y-F, Sha B-Y, Liu X-G, Yan H, et al. Genome-wide copy number variation association study suggested *VPS13B* gene for osteoporosis in Caucasians. Osteoporosis international. 2010;21(4):579-87.

231. Estrada K, Styrkarsdottir U, Evangelou E, Hsu Y-H, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nature genetics. 2012;44(5):491-501.

232. Lebeau G, Miller LC, Tartas M, McAdam R, Laplante I, Badeaux F, et al. Staufen 2 regulates mGluR long-term depression and *MAP1b* mRNA distribution in hippocampal neurons. Learning & Memory. 2011;18(5):314-26.

233. Zhang J, Tu Q, Grosschedl R, Kim MS, Griffin T, Drissi H, et al. Roles of *SATB2* in osteogenic differentiation and bone regeneration. Tissue Engineering Part A. 2011;17(13-14):1767-76.

234. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. Nature genetics. 2008;40(5):609-15.

235. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467(7317):832-8.

236. Smith EN, Chen W, Kähönen M, Kettunen J, Lehtimäki T, Peltonen L, et al. Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study. PLoS Genet. 2010;6: p. e1001094

237. Polašek O, Marušić A, Rotim K, Hayward C, Vitart V, Huffman J, et al. Genome-wide association study of anthropometric traits in Korčula Island, Croatia. Croatian medical journal. 2009;50(1):7-16.

238. Nei M, Kumar S. Molecular evolution and phylogenetics: evolutionary change of DNA Sequences: Oxford University Press; 2000.

239. Rosenberg MS. Evolutionary distance estimation and fidelity of pair wise sequence alignment. BMC bioinformatics. 2005;6(1):102.

240. Sheehan S, Harris K, Song YS. Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. Genetics. 2013;194(3):647-62.

241. Toews DP. Evolution: A Genomic Guide to Bird Population History. Current Biology. 2015;25(11):R465-R7.

242. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. Current Biology. 2015;25(10):1375-80.

243. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493-6.

244. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491(7424):393-8.

245. Schubert M, Jónsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, et al. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. Proceedings of the National Academy of Sciences. 2014;111(52):E5661-E9.

246. Zhao S, Zheng P, Dong S, Zhan X, Wu Q, Guo X, et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nature Genetics. 2013;45(1):67-71.

247. Weir JT, Schluter D. Ice sheets promote speciation in boreal birds. Proceedings of the Royal Society of London B: Biological Sciences. 2004;271(1551):1881-7.

248. Jetz W, Thomas G, Joy J, Hartmann K, Mooers A. The global diversity of birds in space and time. Nature. 2012;491(7424):444-8.

249. Ricklefs RE. Adaptations to cold in bird chicks. Physiology of cold adaptation in birds: Springer; 1989. p. 329-38.

250. Vigo G, Williams M, Brightsmith DJ. Growth of Scarlet Macaw (*Ara macao*) chicks in southeastern Peru. Neotropical Ornithology. 2011;22:143-53.

251. Dobzhansky T. Evolution in the tropics. American Scientist. 1950;38(2):209-21.

252. MacArthur RH. The theory of island biogeography: Princeton University Press; 1967.

253. Schable NA, Faircloth BC, Palmer WE, Carroll JP, Burger L, Brennan LA, et al. Tetranucleotide and dinucleotide microsatellite loci from the northern bobwhite (*Colinus virginianus*). Molecular Ecology Notes. 2004;4(3):415-9.

254. Faircloth BC, Terhune TM, Schable NA, Glenn TC, Palmer WE, Carroll JP. Ten microsatellite loci from Northern Bobwhite (*Colinus virginianus*). Conservation genetics. 2009;10(3):535-8.

255. Wong GK-S, Liu B, Wang J, Zhang Y, Yang X, Zhang Z, et al. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature. 2004;432(7018):717-22.

256. Eshleman JA, Malhi RS, Smith DG. Mitochondrial DNA studies of Native Americans: conceptions and misconceptions of the population prehistory of the Americas. Evolutionary Anthropology: Issues, News, and Reviews. 2003;12(1):7-18.

257. Gilbert MTP, Jenkins DL, Götherstrom A, Naveran N, Sanchez JJ, Hofreiter M, et al. DNA from pre-Clovis human coprolites in Oregon, North America. Science. 2008;320(5877):786-9.

258. Waters MR, Forman SL, Jennings TA, Nordt LC, Driese SG, Feinberg JM, et al. The Buttermilk Creek complex and the origins of Clovis at the Debra L. Friedkin site, Texas. Science. 2011;331(6024):1599-603.

259. Waters MR, Stafford TW, McDonald HG, Gustafson C, Rasmussen M, Cappellini E, et al. Pre-Clovis mastodon hunting 13,800 years ago at the Manis site, Washington. Science. 2011;334(6054):351-3.

260. Alroy J. A multispecies overkill simulation of the end-Pleistocene megafaunal mass extinction. Science. 2001;292(5523):1893-6.

261. Firestone RB, West A, Kennett J, Becker L, Bunch T, Revay Z, et al. Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. Proceedings of the National Academy of Sciences. 2007;104(41):16016-21.

262. Pushkina D, Raia P. Human influence on distribution and extinctions of the late Pleistocene Eurasian megafauna. Journal of Human Evolution. 2008;54(6):769-82.

263. Yokoyama Y, Lambeck K, De Deckker P, Johnston P, Fifield LK. Timing of the Last Glacial Maximum from observed sea-level minima. Nature. 2000;406(6797):713-6.

264. Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, et al. The last glacial maximum. Science. 2009;325(5941):710-4.

265. Redford KH, Robinson JG. The game of choice: patterns of Indian and colonist hunting in the Neotropics. American anthropologist. 1987;89(3):650-67.

266. Jackson HE, Scott SL. The faunal record of the southeastern elite: The implications of economy, social relations, and ideology. Southeastern Archaeology. 1995:103-19.

267. Kricher JC. A neotropical companion: an introduction to the animals, plants, and ecosystems of the New World tropics (2nd Edition): Princeton University Press; 1999.

268. Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science. 2012;338(6104):222-6.

269. Caparroz R, Miyaki CY, Bampi MI, Wajntal A. Analysis of the genetic variability in a sample of the remaining group of Spix's Macaw (*Cyanopsitta spixii, Psittaciformes*: Aves) by DNA fingerprinting. Biological Conservation. 2001;99(3):307-11.

270. Hemmings N, West M, Birkhead T. Causes of hatching failure in endangered birds. Biology letters. 2012:rsbl20120655.

271. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics. 2000;156(1):297-304.

272. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science. 2010;328(5978):636-9.

273. Mayer C, Leese F, Tollrian R. Genome-wide analysis of tandem repeats in Daphnia pulex-a comparative approach. BMC Genomics. 2010;11(1):277.

274. Sánchez CC, Smith TP, Wiedmann RT, Vallejo RL, Salem M, Yao J, et al. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. Bmc Genomics. 2009;10(1):559.

275. Marsden HM, Baskett TS. Annual mortality in a banded bobwhite population. The Journal of Wildlife Management. 1958:414-9.

276. Kabat C, Thompson DR. Wisconsin quail, 1834-1962: Population dynamics and habitat management: Wisconsin Conservation Department; 1963.

277. Speake DW. Ecology and management studies of the bobwhite quail in the Alabama Piedmont. Ph.D. Dissertation, Aburn University, AL; 1967.

278. Folk TH, Holmes RR, Grand JB. Variation in northern bobwhite demography along two temporal scales. Population Ecology. 2007;49(3):211-9.

279. Guthery FS, Lusk JJ. Radiotelemetry studies: are we radio-handicapping northern bobwhites? Wildlife Society Bulletin. 2004;32(1):194-201.

280. Lande R, Engen S, Saether B-E. Stochastic population dynamics in ecology and conservation: Oxford University Press Oxford; 2003.

281. Brightsmith D, Hilburn J, Del Campo A, Boyd J, Frisius M, Frisius R, et al. The use of hand-raised psittacines for reintroduction: a case study of scarlet macaws (Ara macao) in Peru and Costa Rica. Biological Conservation. 2005;121(3):465-72.

282. Vaughan C, Nemeth NM, Cary J, Temple S. Response of a Scarlet Macaw, *Ara macao*, population to conservation practices in Costa Rica. Bird Conservation International. 2005;15(02):119-30.

283. Bouzat JL, Strem RI. Population viability analysis of the blue-throated macaw (*Ara glaucogularis*) using individual-based and cohort-based PVA programs. Open Conservation Biology Journal. 2012;6(1):12.

284. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nature Reviews Genetics. 2013;14(12):827-39.

285. Eo SH, Wares JP, Carroll JP. Subspecies and units for conservation and management of the northern bobwhite in the eastern United States. Conservation genetics. 2010;11(3):867-75.

286. Ryder OA. Species conservation and systematics: the dilemma of subspecies. Trends in Ecology & Evolution. 1986;1(1):9-10.

287. Moritz C. Defining 'evolutionarily significant units' for conservation. Trends in Ecology & Evolution. 1994;9(10):373-5.

288. Avise JC. Phylogeography: the history and formation of species: Harvard university press; 2000.

289. Crandall KA, Bininda-Emonds OR, Mace GM, Wayne RK. Considering evolutionary processes in conservation biology. Trends in ecology & evolution. 2000;15(7):290-5.

290. Zink RM. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. Proceedings of the Royal Society of London B: Biological Sciences. 2004;271(1539):561-4.

291. Esler D, Iverson SA, Rizzolo DJ. Genetic and demographic criteria for defining population units for conservation: the value of clear messages. The Condor. 2006;108(2):480-3.

292. Palsbøll PJ, Berube M, Allendorf FW. Identification of management units using population genetic data. Trends in Ecology & Evolution. 2007;22(1):11-6.

293. Edwards SV. Long-distance gene flow in a cooperative breeder detected in genealogies of mitochondrial DNA sequences. Proceedings of the Royal Society of London B: Biological Sciences. 1993;252(1335):177-85.

294. Templeton AR. The meaning of species and speciation: a genetic perspective. The units of evolution: Essays on the nature of species. 1989:159-83.

295. Templeton AR. The role of molecular genetics in speciation studies. Molecular approaches to ecology and evolution: Springer; 1998. p. 131-56.

296. Avise JC. Molecular markers, natural history and evolution: Springer Science & Business Media; 2012.

297. Smith TB, Wayne RK, Science AAftAo. Molecular genetic approaches in conservation: Oxford University Press New York; 1996.

298. Newton A, Allnutt T, Gillies A, Lowe A, Ennos R. Molecular phylogeography, intraspecific variation and the conservation of tree species. Trends in Ecology & Evolution. 1999;14(4):140-5.

299. Marshall HD, Baker AJ. Structural conservation and variation in the mitochondrial control region of fringilline finches (*Fringilla spp*.) and the Greenfinch (*Carduelis chloris*). Molecular Biology and Evolution. 1997;14(2):173-84.

300. Baker A, Marshall H. Mitochondrial control region sequences as tools for understanding evolution. Avian molecular evolution and systematics. 1997:51-82.

301. Simon C. Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. Molecular techniques in taxonomy: Springer; 1991. p. 33-71.

302. Li W-H, Wu C-I, Luo C-C. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. Journal of Molecular Evolution. 1984;21(1):58-71.

303. Hillis DM, Davis SK. Evolution of ribosomal DNA: fifty million years of recorded history in the frog genus Rana. Evolution. 1986:1275-88.

304. Wenink PW, Baker AJ, Tilanus M. Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpina*). Proceedings of the National Academy of Sciences. 1993;90(1):94-8.

305. Wenink PW, Baker AJ, Rosner H-U, Tilanus MG. Global mitochondrial DNA phylogeography of holarctic breeding dunlins (*Calidris alpina*). Evolution. 1996:318-30.

306. Quinn T. The genetic legacy of Mother Goose–phylogeographic patterns of lesser snow goose, *Chen caerulescens caerulescens*, maternal lineages. Molecular Ecology. 1992;1(2):105-17.

307. Rubinoff D. Utility of mitochondrial DNA barcodes in species conservation. Conservation Biology. 2006;20(4):1026-33.

308. Walker A, Smith S, Smith S. Mitochondrial DNA and human evolution. Nature. 1987;325:1-5.

309. Carr SM, Brothers AJ, Wilson AC. Evolutionary inferences from restriction maps of mitochondrial DNA from nine taxa of Xenopus frogs. Evolution. 1987:176-88.

310. Clary DO, Wolstenholme DR. Drosophila mitochondrial DNA: conserved sequences in the A+ T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. Journal of molecular evolution. 1987;25(2):116-25.

311. Uhlenbusch I, McCracken A, Gellissen G. The gene for the large (16S) ribosomal RNA from the *Locusta migratoria* mitochondrial genome. Current genetics. 1987;11(8):631-8.

312. Haucke H-R, Gellissen G. Different mitochondrial gene orders among insects: exchanged tRNA gene positions in the *COII*/*COIII* region between an orthopteran and a dipteran species. Current genetics. 1988;14(5):471-6.

313. Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, et al. A draft de novo genome assembly for the northern bobwhite (*Colinus virginianus*) reveals evidence for a rapid decline in effective population size beginning in the Late Pleistocene. PLoS ONE. 2014;9(3).

314. Meiklejohn KA, Danielson MJ, Faircloth BC, Glenn TC, Braun EL, Kimball RT. Incongruence among different mitochondrial regions: A case study using complete mitogenomes. Molecular phylogenetics and evolution. 2014;78:314-23.

315. Mindell DP, Sorenson MD, Dimcheff DE. An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. Molecular Biology and Evolution. 1998;15:1568-71.

316. García-Trejo EA, De Los Monteros AE, Arizmendi MDC, Navarro-Siüenza AG. Molecular systematics of the red-bellied and golden-fronted woodpeckers. The Condor. 2009;111(3):442-52.

317. Cibois A, Thibault J-C, Bonillo C, Filardi CE, Watling D, Pasquet E. Phylogeny and biogeography of the fruit doves (Aves: *Columbidae*). Molecular phylogenetics and evolution. 2014;70:442-53.

318. Labuschagne C, Kotzé A, Grobler JP, Dalton DL. The complete sequence of the mitochondrial genome of the African Penguin (*Spheniscus demersus*). Gene. 2014;534(1):113-8.

319. Nishibori M, Hanazono M, Yamamoto Y, Tsudzuki M, Yasue H. Complete nucleotide sequence of mitochondrial DNA in White Leghorn and White Plymouth Rock chickens. Animal Science Journal. 2003;74(5):437-9.

320. Guan X, Silva P, Gyenai KB, Xu J, Geng T, Tu Z, et al. The mitochondrial genome sequence and molecular phylogeny of the turkey, Meleagris gallopavo. Animal genetics. 2009;40(2):134-41.

321. Slack KE, Janke A, Penny D, Arnason U. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene. 2003;302(1):43-52.

322. Nishibori M, Hayashi T, Tsudzuki M, Yamamoto Y, Yasue H. Complete sequence of the Japanese quail (*Coturnix japonica*) mitochondrial genome and its genetic relationship with related species. Animal genetics. 2001;32(6):380-5.

323. Pereira SL, Baker AJ. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. BMC Evolutionary Biology. 2004;4(1):17.

324. Li M, Schönberg A, Schaefer M, Schroeder R, Nasidze I, Stoneking M. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. The American Journal of Human Genetics. 2010;87(2):237-49.

325. Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, et al. Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. Genome Biology. 2011;12(6):R59.

326. Mundy N, Winchell C, Woodruff D. Tandem repeats and heteroplasmy in the mitochondrial DNA control region of the loggerhead shrike (*Lanius ludovicianus*). Journal of Heredity. 1996;87(1):21-6.

327. Kvist L, Martens J, Nazarenko AA, Orell M. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Molecular Biology and Evolution. 2003;20(2):243-7.

328. Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Molecular Biology and Evolution. 2007;24(1):269-80.

329. Lopes I, Tomasulo-Seccomandi A, Bryan Jr A, Brisbin Jr I, Glenn T, Del Lama S. Genetic status of the wood stork (*Mycteria americana*) from the southeastern

United States and the Brazilian Pantanal as revealed by mitochondrial DNA analysis. Genetics and Molecular Research. 2011;10(3):1910.

330. Sammler S, Bleidorn C, Tiedemann R. Full mitochondrial genome sequences of two endemic Philippine hornbill species (Aves: *Bucerotidae*) provide evidence for pervasive mitochondrial DNA recombination. BMC genomics. 2011;12(1):35.

331. He X-L, Ding C-Q, Han J-L. Lack of structural variation but extensive length polymorphisms and heteroplasmic length variations in the mitochondrial DNA control region of highly inbred Crested Ibis, *Nipponia nippon*. PLoS ONE. 2013; 8:e66324.

332. Smigrodzki RM, Khan SM. Mitochondrial microheteroplasmy and a theory of aging and age-related disease. Rejuvenation research. 2005;8(3):172-98.

333. Lin MT, Simon DK, Ahn CH, Kim LM, Beal MF. High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain. Human molecular genetics. 2002;11(2):133-45.

334. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. Nature. 2010;464(7288):610-4.

335. Sondheimer N, Glatz CE, Tirone JE, Deardorff MA, Krieger AM, Hakonarson H. Neutral mitochondrial heteroplasmy and the influence of aging. Human molecular genetics. 2011;20(8):1653-9.

336. Royston SR, Carr SM. Conservation genetics of high-arctic Gull species at risk: I. Diversity in the mtDNA control region of circumpolar populations of the Endangered Ivory Gull (*Pagophila eburnea*). Mitochondrial DNA. 2014(0):1-5.

337. Hailer F, James HF, Olson SL, Fleischer RC. Distinct and extinct: Genetic differentiation of the Hawaiian eagle. Molecular phylogenetics and evolution. 2015;83:40-3.

338. Ferrando A, Manunza A, Jordana J, Capote J, Pons A, Pais J, et al. A mitochondrial analysis reveals distinct founder effect signatures in Canarian and Balearic goats. Animal genetics. 2015;46(4):452-6.

339. Heikkinen M, Ruokonen M, Alexander M, Aspi J, Pyhäjärvi T, Searle J. Relationship between wild greylag and European domestic geese based on mitochondrial DNA. Animal genetics. 2015.

340. Dudu A, Georgescu SE, Costache M. Molecular analysis of phylogeographic subspecies in three Ponto-Caspian sturgeon species. Genetics and molecular biology. 2014;37(3):587-97.

341. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution. 1999;16(1):37-48.

342. Hudson R, Boos DD, Kaplan N. A statistical test for detecting geographic subdivision. Molecular biology and evolution. 1992;9(1):138-51.

343. Nei M. Molecular evolutionary genetics: Columbia university press; 1987.

344. Harpending HC, Sherry ST, Rogers AR, Stoneking M. The genetic structure of ancient human populations. Current Anthropology. 1993:483-96.

345. Vinas J, Bremer JA, Pla C. Phylogeography of the Atlantic bonito (*Sarda sarda*) in the northern Mediterranean: the combined effects of historical vicariance, population expansion, secondary invasion, and isolation by distance. Molecular phylogenetics and evolution. 2004;33(1):32-42.

346. Horne JB, van Herwerden L, Choat JH, Robertson D. High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. Molecular phylogenetics and evolution. 2008;49(2):629-38.

347. Ray N, Currat M, Excoffier L. Intra-deme molecular diversity in spatially expanding populations. Molecular Biology and Evolution. 2003;20(1):76-86.

348. Aris-Brosou S, Excoffier L. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Molecular Biology and Evolution. 1996;13(3):494-504.

349. Marjoram P, Donnelly P. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics. 1994;136(2):673-83.

350. Chen B, Pedro P, Harbach R, Somboon P, Walton C, Butlin R. Mitochondrial DNA variation in the malaria vector *Anopheles minimus* across China, Thailand and Vietnam: evolutionary hypothesis, population structure and population history. Heredity. 2011;106(2):241-52.

351. Inoue K, Monroe E, Elderkin C, Berg D. Phylogeographic and population genetic analyses reveal Pleistocene isolation followed by high gene flow in a wide ranging, but endangered, freshwater mussel. Heredity. 2014;112(3):282-90.

352. Weckstein JD, Afton AD, Zink RM, Alisauskas RT. Hybridization and population subdivision within and between Ross's Geese and Lesser Snow Geese: a molecular perspective. The Condor. 2002;104(2):432-6.

353. Johnsgard PA. A summary of intergeneric new world quail hybrids, and a new intergeneric hybrid combination. The Condor. 1970:85-8.

354. Shupe TE. Frequency of northern bobwhite x scaled quail hybridization. The Wilson bulletin. 1990;102(2):352-3.

355. Wang N, Kimball RT, Braun EL, Liang B, Zhang Z. Assessing phylogenetic relationships among Galliformes: a multigene phylogeny with expanded taxon sampling in Phasianidae. 2013.

356. Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 1993;135(2):599-607.

357. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. Molecular biology and evolution. 1992;9(3):552-69.

358. Watterson G. On the number of segregating sites in genetical models without recombination. Theoretical population biology. 1975;7(2):256-76.

359. Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics. 1991;129(2):555-62.

360. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585-95.

361. Fu Y-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997;147(2):915-25.

362. Egea R, Casillas S, Barbadilla A. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic acids research. 2008;36(suppl 2):W157-W62.

363. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351(6328):652-4.

364. Smith NG, Eyre-Walker A. Adaptive protein evolution in Drosophila. Nature. 2002;415(6875):1022-4.

365. Eyre-Walker A. Changing effective population size and the McDonald-Kreitman test. Genetics. 2002;162(4):2017-24.

366. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688-90.

367. Tajima F. The effect of change in population size on DNA polymorphism. Genetics. 1989;123(3):597-601.

368. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25(11):1451-2.

369. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution. 1993;10(3):512-26.

370. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources. 2010;10(3):564-7.

371. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution. 2013;30(12):2725-9.

372. Evans K, Burger Jr L, Palmer W. Genetic structure of northern bobwhites in northeast Mississippi and. 2012.

373. Fies M, Puckett K, Larson-Brogdon B, editors. Breeding season movements and dispersal of northern bobwhites in fragmented habitats of Virginia. Proceedings of the National Quail Symposium; 2002.

374. Liberati MR. Spring Dispersal and Breeding Ecology of Northern Bobwhite in Southwest Ohio: The Ohio State University; 2013.

375. Howard WE. Innate and environmental dispersal of individual vertebrates. American Midland Naturalist. 1960:152-61.

376. Greenwood PJ, Harvey PH. The natal and breeding dispersal of birds. Annual review of ecology and systematics. 1982:1-21.

377. Walters JR. Dispersal behavior: an ornithological frontier. The Condor. 2000;102(3):479-81.

APPENDIX A

Table S11-A. Bobwhite Quail *De Novo* Outlier Contigs (NB1.0) from a Genome-wide Analysis of Divergence with the Chicken.

Bobwhite Quail Contig[1]	Outlier Direction[2]	Predicted Content[3]	Predicted Description[4]
42497	Conserved	NRG2	(I) Neuregulin 2
1920	Conserved	*CFDP1*	(I,E) Craniofacial Development Protein 1[9]
28277	Conserved	*PDZD2*	(I,E) PDZ Domain-containing Protein 2
18853	Conserved	KIAA1328	(I,E) Hinderin
35615	Conserved	Noncoding	Between *GALT* and *CNTRF* [5]
13242	Conserved	*TLN1*	(I,E) Talin-1
17476	Conserved	*LDB2*[8]	(I,E) LIM Domain-Binding Protein 2
66729	Conserved	*WDR7*	(I,E) WD Repeat-Containing 7
37232	Conserved	Noncoding	Between Shugoshin-like and *ZNF385D*[5]
2112	Conserved	Noncoding	Between *TNC* and *PAPPA*[5]
78581	Conserved	*BCL11B*	(I) B-cell CLL/lymphoma 11B
17775	Conserved	*CDH4*	(I) Cadherin-4
27080	Conserved	*ZNF521*[8]	(I) Zinc Finger Protein 521
30465	Conserved	*FBXL8*	(I,E) F-box and Leucine-Rich Repeat Protein 8
15192	Conserved	*PRDM11*	(I,E) PR Domain Containing 11
3911	Conserved	*SUFU*	(I,E) Suppressor of Fused Homolog
19454	Conserved	*LTBP2*	(I,E) Latent Transforming Growth Factor Beta Binding Protein 2
7216	Conserved	*GRM3*	(I,E) Glutamate Receptor, Metabotropic 3[9]
93461	Conserved	*TAOK1*	(I,E) Serine/Threonine-protein Kinase Tao1
47909	Conserved	*VPS13B*	(I,E) Vacuolar Protein Sorting 13 Homolog B
1508	Conserved	*EPHA5*	(I,E) Ephrin Type-A Receptor 5
64848	Conserved	*SSBP2*	(I,E) Single-Stranded DNA-Binding Protein 2
14294	Conserved	Noncoding[8]	Between *TBC1D5* and *SATB1*[5]
1560	Conserved	*MAMLD1*	(I,E) Mastermind-like Domain Containing 1

20910	Conserved	Noncoding[8]	Between *LPP*[9] and *BCL6*[5]
25091	Conserved	*GALK2*	(I,E) N-Acetylgalactosamine Kinase
14164	Conserved	*SLC28A2*	(I,E) Nucleoside Cotransporter 2
14164	Conserved	*ALPK3*	(I,E) Alpha-Protein Kinase 3[9]
19319	Conserved	*CADPS*	(I,E) Calcium-Dependent Secretion Activator 1[9]
17011	Conserved	*NFASC*	(I,E) Neurofascin
4036	Conserved	*PTPRF*	(I,E) Protein Tyrosine Phosphatase, Receptor Type, F
106766	Conserved	Noncoding	Between *CUZD1* and Uncharacterized Protein Loc769645[6]
1952	Conserved	*VSX2*[8]	(I,E) Visual System Homeobx 2
1952	Conserved	*ACSS1*	(I,E) Acyl-coa Synthetase Short-chain Family Member 1-like
56418	Conserved	*TBC1D5*	(I,E) TBC1 Domain Family Member 5
69832	Conserved	Noncoding	Between *NCKAP5* and Aplha-1,6-mannosylgycoprotein 6-beta-N- acetylglucosaminyl Transferase[5]
42835	Conserved	Noncoding	Between Protocadherin-19 Precursor and *DIAPH1*[6]
13148	Conserved	*MEIS2*	(I,E) Homeobox Protein MEis2[9]
15075	Conserved	*CELF4*	(I,E) CGBP Elay-like Family Member 4
28121	Conserved	*RELN*	(I) Reelin
70673	Conserved	*SETBP1*	(I) Set-Binding Protein 1
53420	Conserved	*ZNF652*	(I,E) Zinc Finger Protein 652
51920	Conserved	*TRMT61A*	(I,E) tRNA (adenine-N(1)-methyltransferase Catalytic Subunit TRMT61A-like
14518	Conserved	*ELP4*	(I,E) Elongator Complex Protein 4
2537	Conserved	Noncoding	Between *KCNJ2* and *SOX9*[5, 9]
63983	Conserved	*FOXP1*	(I,E) Forkhead Box Protein P1
5280	Conserved	Noncoding	Between *GATA5* and *SLCO4A1*[5]
69017	Conserved	*CELF4*	(I) CUGBP, Ekave-like Member 4
1277	Conserved	LOC776265	(I,E) Uncharacterized Protein Loc776265
57732	Conserved	*CUX1*	(I,E) Cut-like Homeobox 1
5181	Conserved	Noncoding	Between *IGSF11* and *LSAMP*[5]
39333	Conserved	Noncoding	Between *CENT3* and *MEF2C*[5, 9]
8712	Conserved	Noncoding	Between *BMF* and LOC729466[5]

53126	Conserved	Noncoding	Between Uncharacterized Protein Loc100857170 and *TMEM121*[6]
75545	Conserved	*FIGN*	(I,E) Fidgetin
3614	Conserved	Noncoding	Between *TRIM66* and Rhomobotin-1-like[6]
2456	Conserved	LOC427016	(I,E) Rho-guanine Nucleotide Exchange Factor-like
40285	Conserved	Noncoding	Between *VGLL3* and *CADM2*[6]
43566	Conserved	Noncoding	Before *PLA2G4A*[6]
83177	Conserved	*EBF3*	(I,E) Transcription Factor COE3
3267	Conserved	Noncoding	Between *ENC1* and Rho-guanine Nucleotide Exchange Factor-like[5]
1577	Conserved	Noncoding	Between *FIGN3* and Uncharacterized protein LOC100858207[6]
59785	Conserved	Noncoding	Between *BCA2* and *SOX5*[5]
5671	Conserved	*ATP10B*	(I,E) Phospholipid-transporting ATPase VB
23853	Conserved	Noncoding	Between *CHIC2* and *LNX1*[5]
40758	Conserved	*NFIB*	(I) Nuclear Factor 1B-type
4309	Conserved	Noncoding	Between *LPL*[9] and *PSD3*[5]
2492	Conserved	Noncoding	Between *CEIF4* and *FAF1*[5]
44303	Conserved	Noncoding	Between *IGSF11* and *LSAMP*[5]
12224	Conserved	*HIF1A*	(I,E) Hypoxia-inducible Factor 1-alpha
30481	Conserved	*RORB*	(I) Nuclear Receptor ROR-beta
41810	Conserved	Noncoding	Between *NEDD1* and *TMPO*[5]
31345	Conserved	Noncoding	Between *SMYD2* and *PROX1*[5]
27458	Conserved	*EFNA5*	(I,E) Ephrin-A5 Precursor
4250	Conserved	*NTRK3*	(I,E) NT-3 Growth Factor Receptor Precursor
30943	Conserved	Noncoding	Between *LRRC28* and *MEF2A*[6, 9]
202620	Diverged	Noncoding	No Repeats, Unknown Orthology
19925	Diverged	Noncoding	No Repeats, Unknown Orthology
226794	Diverged	Noncoding	No Repeats, Unknown Orthology
136209	Diverged	Noncoding	No Repeats, Unknown Orthology
160937	Diverged	Noncoding	No Repeats, Between Acetylglucosaminyltransferase-like and *EPHA1* in *X. tropicalis* (NW_003808088.1), 84% ID Across 44bp

198701	Diverged	Noncoding	No Repeats, Between *RIF1* and *ARL5A* in *O. anatinus* (NW_001794453.1), 89% ID Across 38bp
109004	Diverged	Noncoding	No Repeats, Unknown Orthology
109025	Diverged	Noncoding	No Repeats, Best Hit to *C. virginianus* DNA for Female-specific 0.4 kb BamHI Repetitive Unit, 72% ID Across 104bp
242738	Diverged	Noncoding	No Repeats, Between *RFK* and *GCNT* in *O. cuniculus* (NW_003159226.1), 84% ID Across 45bp
233571	Diverged	Noncoding	High Repeats, Top Hit to ZF ChrUn (NW_002218881.1), 68% ID Across 397bp
215237	Diverged	Noncoding	No Repeats, Unknown Orthology
241371	Diverged	Noncoding	No Repeats, Unknown Orthology, Best Hit *D. novemcinctus* (NW_004461987.1), 79% ID Across 58bp
269471	Diverged	Noncoding	No Repeats, Unknown Orthology
266775	Diverged	Noncoding	No Repeats, Unknown Orthology
255170[7]	Diverged	*CCNL2*	(E) No Repeats, Short Hit to *S. boliviensis* (XM_003939658.1), 91% ID Across 33bp
285736	Diverged	Noncoding	No Repeats, Short Hit to *C.. jacchus* Chromosome 19 (NW_003184577.1), Unknown Orthology
274292[7]	Diverged	*TNIK*	(I) No Repeats, Best Hit to *P. Abelii* (NW_002877893.1), 81% ID Across 53bp
286938	Diverged	Noncoding	No Repeats, Between *SEL1L* and Sodium Dependent Phosphate Transport Protein 2B in *F. catus* (NC_018726.1), 93% ID Across 30bp
295704	Diverged	Noncoding	No Repeats, Between *FTMT* and *PRR16* in *P. paniscus* (NW_003870563.1), 89% ID Across 37bp
301824	Diverged	Noncoding	No Repeats, Between T-Cell Ecto-ADP- ribosyltranserase 2-like and CLPX in *C. porcellus*, 96% ID Across 28bp
277525[7]	Diverged	*PARD3*	(I) No Repeats, Unknown Orthology, Short hit to Walrus (NW_004451169.1), 81% ID Across 53bp
299763	Diverged	Noncoding	No Repeats, Unknown Orthology
315442	Diverged	Noncoding	No Repeats, Unknown Orthology
318100	Diverged	Noncoding	No Repeats, Between Inducible T-Cell Co-stimulator-like and *NRP2* in horse (NC_009161.2), 85% ID Across 41bp
335973	Diverged	Noncoding	No Repeats, Unknown Orthology
333349	Diverged	Noncoding	No Repeats, Between *IGFBP7* and *LPHN3* in *M. musculus* (NC_000071.6), 81% ID Across 35bp
335787	Diverged	Noncoding	No Repeats, Unknown Orthology
336370	Diverged	Noncoding	No Repeats, Unknown Orthology
337331[7]	Diverged	*PSMB8*	(E) No Repeats, Unknown Orthology, *G. cirratum* Clone (AC165195.3), 79% ID Across 30bp
343316	Diverged	Unknown	No Repeats, Unknown Orthology, Short hit to *A. carolinensis* k26:28253705 Transcribed RNA

			Sequence, 72% ID Across 62bp
343903	Diverged	Noncoding	No Repeats, Unknown Orthology
350705	Diverged	Noncoding	No Repeats, Unknown Orthology
356261	Diverged	Noncoding	No Repeats, Unknown Orthology, Top Hit *A. carolinensis* k26:15424195 Transcribed RNA Sequence (GAGG010186469.10), 80% ID Across 54bp
35292	Diverged	Noncoding	No Repeats, Unknown Orthology
357039	Diverged	Noncoding	No Repeats, Unknown Orthology
356932	Diverged	Noncoding	No Repeats, Between Leydig Cell Tumor 10 kDa Protein-like and LOC100996639, NC_005105.3
366276	Diverged	Noncoding	No Repeats, Unknown Orthology
364995	Diverged	Noncoding	No Repeats, In Assembly Gap of NW_004504331.1, Unknown Orthology
354189	Diverged	Noncoding	No Repeats, Unknown Orthology
367189	Diverged	Noncoding	No Repeats, Unknown Orthology
368722[7]	Diverged	*KRT26*	(I) No Repeats, Unknown Orthology, Short Hit to Rabbit (NW_003159313.1), 89% ID Across 35bp
363506[7]	Diverged	*BRF1*	(I) No Repeats, Unknown Orthology, Short Hit to *Homo sapiens* (NG_029489.1), 89% ID Across 35bp
369836[7]	Diverged	LOC10095172	(I) No Repeats, Unknown Orthology, Short Hit to Embigin-like in *O. garnettii* (NW_003852400.1), 93% ID Across 29bp
370861	Diverged	Noncoding	No Repeats, Beside *EIF4ENIF* in *T. manatus* (NW_004443997.1), 80% ID Across 51bp
373008[7]	Diverged	*PPAPDC1A*	(I) No Repeat, Short Hit to Orca Unplaced scaffold (NW_004438429.1), 91% ID Across 34bp
373159	Diverged	Noncoding	No Repeats, Unknown Orthology, Short Hit to Hamster Unplaced Scaffold (NW_003614382.1), 85% ID Across 41bp
374055	Diverged	Noncoding	No Repeats, Unknown Orthology, Short Hit to *A. nancymaae* (NT_165745), 94%ID Across 32bp
10512[7]	Diverged	*DENND5A*	(I) No Repeats, Unknown Orthology, Short hit to *S. harrisii*, (NW_003846890.1), 88% ID Across 39bp
152990[7]	Diverged	*PKD2*	(I) No, Repeats, Unknown Orthology, Short hit to Elephant Unplaced scaffold (NW_003573450), 85% ID Across 39bp
105451	Diverged	Unknown	No Repeats, Unknown Orthology, Short Hit to Walrus Unplaced Scaffold (NW_004450309.1), 87% ID Across 38bp
217228	Diverged	Noncoding	No Repeats, Between *TRSPS1* and *CSMD3* in *O. garnettii* (NW_003852399.1), 83% ID Across 47bp
311181[7]	Diverged	*CSMD2*	(I) No Repeats, Unknown Orthology, Best Hit to Walrus Unplaced Scaffold (NW_004451812), 92% ID Across 36bp

332175	Diverged	Noncoding	No Repeats, Unknown Orthology, Best Hit to Bee Unplaced Scaffold (NW_003797141.1), 89% ID 35bp
72085	Diverged	Noncoding	No Repeats, Unknown Orthology, Best Hit Between LOC100856132 and *LACC1* in Wolf (NC_006604.3)

[1] NB1.0 simple *de novo* contig ID.
[2] The direction of the outlier in the full blastn distribution for the comparative genome alignment with chicken.
[3] Concise prediction (top blastn hit) of the genomic information content for each contig (gene symbol, noncoding, or unknown).
[4] Detailed description of the genomic information content for each contig, as evidenced by blastn searches of refseq_genomic, reseq_rna, and nr/nt, with repeat content predicted by RepeatMasker. Outliers for conservation were annotated based on the Chicken Genome. ZF indicates *Taeniopygia guttata*. (I) indicates intron(s), (E) indicates exon(s), and (I, E) indicates both. Note, the blast databases are dynamic, and therefore, descriptions correspond to results achieved at the time of analysis (Chicken 4.0 and Zebra Finch Build 3.2.4).
[5] Genes are predicted to be syntenic and proximal in both the chicken and zebra finch genomes via blastn and/or NCBI Map Viewer.
[6] Synteny and proximity of genes could not be conclusively determined using the chicken and zebra finch genome resources.
[7] Used corresponding scaffolds to confirm the predicted intron sequence.
[8] Gene/genomic region was found to also be an outlier in the scarlet macaw genome analysis (Seabury et al. 2013).
[9] Previously reported to be under purifying selection (see references).

Table S11-B. Bobwhite Quail *De Novo* Outlier Contigs (NB1.0) from a Genome-wide Analysis of Divergence with the Zebra Finch.

Bobwhite Quail Contig[1]	Outlier Direction[2]	Predicted Content[3]	Predicted Description[4]
13159	Conserved	*ZFHX4*	(I,E) Zinc Finger Homeobox Protein 4
51931	Conserved	*SDCCAG8*	(I,E) Serologically Defined Colon Cancer Antigen 8 Homolog
25364	Conserved	*TENM1*	(I,E) Terneurin-1like
61326	Conserved	Noncoding	No Repeats
136	Conserved	Mitochondria[8]	Complete Annotated Genome (13 Protein Coding Genes, 21 tRNA Genes, 2 rRNA Genes)[9]
10989	Conserved	*SACS*[8]	(I,E) Sacsin
80615	Conserved	Noncoding	No Repeats
36366	Conserved	*ZFHX4*	(I) Zinc finger Homeobox Protein 4
10052	Conserved	*BMPR2*[8]	(I,E) Bone Morphogenetic Protein Receptor, Type II[9]
79692	Conserved	*CYP7B1*	(I,E) 25-hydoxycholesterol 7-alpha-hydroxylase-like
22640	Conserved	*C7ORF10*[8]	(I) CaiB/baiF CoA-transferase Family Protien C7orf10 Homolog
47261	Conserved	*ARHGEF38*	(I,E) Rho Guanine Nucleotide Exchange Factor (GEF) 38
37709	Conserved	*SOX5*	(I,E) Transcription Factor SOX-5
42610	Conserved	Noncoding	No Repeats
14488	Conserved	Noncoding	Between *MEIS1* and *ETAA1*[6]
4881	Conserved	*ST6GALNAC3*	(I) Alpha-N-acetylgalactosaminide Alpha-2,6-sialyltransferase 3
159316	Conserved	*ZAK*	(I,E) Mitogen-activated Protein Kinase MLT-like
32397	Conserved	*JAG1*	(I,E) Low Quality Protein: Protein Jagged-1[9]
58279	Conserved	Noncoding[8]	Between *GPATCH2* and *ESRRB*[6]
91360	Conserved	*VPS13B*	(I,E) Vacuolar Protein Sorting-associated Protein 13B
121848	Conserved	*SEMA3A*[8]	(I,E) Semaphorin 3A
82559	Conserved	*VTIIA*[8]	(I,E) Vesicle Transport Through Interaction with t-SNAREs Homolog 1A
16479	Conserved	Noncoding	Between *CDH2* and *DSC1*[6]
15322	Conserved	*BRSK2*[8]	(I,E) Serine/Threonine-protein Kinase BrSK2
123692	Conserved	*HIC2*	(I,E) Hypermethylated in cancer protein

89827	Conserved	*CCDC88C*	(I,E) Protein Dapple
561	Conserved	Noncoding	Between *APOB*[9] and *KLHL29*
12390	Conserved	*CAMK2G*	(I,E) Calcium/calmodulin-dependent Protein Kinase Type II Subunit Gamma
137453	Conserved	*KIF26A*	(I,E) Kinesin-like Protein KIF26A
26513	Conserved	*AKAP6*	(I,E) A-kinase Anchor Protein 6
3815	Conserved	*URI1*	(I,E) Unconventional Prefolding RPB5 Interactor 1
10346	Conserved	*GRIA1*	(I,E) Glutamate Receptor 1
59810	Conserved	Noncoding	Between Serine/threonine-protein Kinase and LIM Domain Only 4 (*LMO4*)
8327	Conserved	*TJAP1*	(I,E) Tight Junction-associated Protein 1
85351	Conserved	Noncoding[8]	Between *TP63* and *LPP*[5, 9]
24959	Conserved	Noncoding[8]	Between *STX16* and *APCDD1L*[5]
59816	Conserved	Noncoding	Between *ADORA2A* and *UPB1*[5]
37555	Conserved	Noncoding	Between *MCTP2* and *COUOP*[6]
12195	Conserved	Noncoding	Between *KLF5* and *KLF12*[5]
36940	Conserved	*STAU2*	(I,E) Double-stranded RNA-binding Protein Staufen Homolog 2
28591	Conserved	*BTRC*	(I,E) F-box/WD Repeat-containing Protein 1A
63170	Conserved	*GMDS*	(I,E) GDP-mannose 4,6 Dehydratase
41774	Conserved	LOC101232932	(I,E) Uncharacterized LOC101232932
28631	Conserved	Noncoding	Between *MOXD1* and *WISP2*[6]
3326	Conserved	*PHF21A*	(I,E) PHD Finger Protein 21A
23126	Conserved	*GJA1*	(I,E) Gap Junction Alpha-1 Protein[9]
19681	Conserved	*LHX9*	(I,E) Lim Homeobox 9
60440	Conserved	Noncoding	Between *CDH13* and *MPHOSPH6*[5]
91692	Conserved	Noncoding	Succeeding *RBMS1*[6]
7528	Conserved	*PAX2*[8]	(I,E) Paired Box Protein Pax2-A[9]
141786	Conserved	Noncoding	Between *TOX3* and *SALL1*[6]
68750	Conserved	Noncoding[8]	Between *ALCAM* and *ZPLD1*[6]
12466	Conserved	Noncoding	Between *BNC2* and *CCDC171*[6]
78581	Conserved	*BCL11B*	(I) B-cell Lymphoma/Leukemia 11B

109984	Conserved	*SATB2*[8]	(I,E) DNA-binding Protein SATB2
9683	Conserved	*RAD51B*	(I) DNA Repair Protein RAD51 Homolog 2[9]
11350	Conserved	Noncoding	Between *RORA* and *NARG2*[5, 9]
68805	Conserved	*FHOD3*	(I,E) FH1/FH2 Domain-containing Protein 3
64455	Conserved	Noncoding	Between *BARHL2* and BIRD Complex Subunit *ZNF36*[6]
122037	Conserved	*TRPS1*	(I,E) Zinc Finger Transcription Factor Trps1
18565	Conserved	*MYLK4*	(I,E) Myosin Light Chain Kinase, Smooth Muscle-like
36442	Conserved	*BTBD11*	(I,E) Ankyrin Repeat and BTB/POZ Domain-containing Protein BTBD11
9672	Conserved	Noncoding	Between *KLF5* and *KLF12*[5]
15192	Conserved	*ZNF862*	(I,E) Zinc Finger Protein 862
28210	Conserved	Noncoding	Between *EBF1* and *CLINT1*[6]
93289	Conserved	C10ORF11	(I) Leucine-rich Repeat-containing Protein C10orf11 Homolog
27080	Conserved	*ZNF521*[8]	(I,E) Zinc Finger Protein 521
6375	Conserved	*PLCB4*	(I,E) Phospholipase C, Beta 4
13267	Conserved	Noncoding	Between *USP25* and *CXADR*[6]
117769	Conserved	*PODXL2*	(I,E) Podocalyxin-like Protein 2
122488	Conserved	Noncoding	Between *ACPL2* and *EPHA4*[5]
144431	Conserved	Noncoding	Between *EYA1* and *MSC*[5]
3729	Conserved	*PTPRZ1*	(I,E) Receptor-type Tyrosine-protein Phosphatase Zeta Precursor
71045	Conserved	*GJB7*	(I,E) Gap Junction Beta-7 Protein
71045	Conserved	*ZNF292*	(I,E) Zinc Finger Protein 292
109025	Diverged	*BAMHI*	No Repeats, *C. virginianus* DNA for Female-specific 0.4 kb BamHI Repetitive Unit, 72% ID Across 104bp
121215	Diverged	Unknown	No Repeats, Short Hit to *Gallus_gallus*-4.0 ChrUn_7180000979433
144064	Diverged	Noncoding	No Repeats, Flanked by *MMP14* in GG (ref|NW_003779907.1|), 72% ID Across 122bp[5]
149824	Diverged	Noncoding	No Repeats, Unknown Orthology
176346	Diverged	Noncoding	No Repeats, Between Neuronal PAS Domain-containing Protein 2-like and *VMA21* in Pekin duck (NW_004677091.1), 81% ID Across 52bp
179703	Diverged	Noncoding	No Repeats, Between *EXOC3L4* and *CDC42BPB* in *O. degus* (NW_004524802.1), 78% ID Across

			59bp
190404	Diverged	Unknown	No Repeats, Short Hit to GG ChrUn_7180000968132 (NW_003771446.1)
193728	Diverged	Noncoding	No Repeats, Unknown Orthology
234157	Diverged	Noncoding	No Repeats, Best Hit to GG ChrUn_7180000977381 (NW_003778581.1), 89% Across 338 bp
234931	Diverged	Noncoding	No Repeats, Unknown Orthology
244978[7]	Diverged	*RNF128*	(I) Moderate Repeats, Best Hit to intron of *RNF128* in GG (NC_006091.3), 72% ID Across 350bp
250106[7]	Diverged	*LDB2*[8]	(I) No Repeats, Short Hit to Intron of *LDB2* in GG (NW_001471685.2), 91% ID Across 32bp[5]
252196	Diverged	Unknown	No Repeats, Unknown Orthology, *O. degus* mRNA (XM_004623416.1), 83% ID Across 46bp
263045	Diverged	Noncoding	No Repeats, Unknown Orthology
278585	Diverged	Noncoding	No Repeats, Between *NR4A2* and G Protein-activated Inward Rectifier Potassium Channel 1-Like in elephant (LOC100668323), NW_003573423.1, 90% ID Across 39bp
296881	Diverged	Noncoding	No Repeats, Between *KCNJ16* and (I,E) *KCNJ2* in GG (NW_004504323.10), 80% ID Across 352bp[5]
299790	Diverged	Noncoding	No Repeats, Between *DDB1* and *CUL4* Associated Factor 13-like and S-phase Kinase-Associated Protein 1-like in *J. jaculus* (LOC101600200), ref|NW_002198637.1, 87%ID across 38bp
311205	Diverged	Noncoding	High Repeats, Top hit to GG ChrW_random_7180000979747, 74% ID Across 242bp
312077	Diverged	Noncoding	No Repeats, Beside EPHB4 on GG (NW_003772415.1), 72% ID Across 306 bp[5]
318709[7]	Diverged	*NGDN*	(I) No Repeats, Unknown Orthology, *S. harrisii* Unplaced Scaffold (NW_003816506.1), 89% ID Across 35bp
320523[7]	Diverged	*ZCCHC2*	(I) No Repeats, Best hit to *G. gallus* (NC_006089.3)
326623	Diverged	Noncoding	No Repeats, GG ChrUn_7180000967474 (NW_003770987.1), 66% ID Across 336bp
326982	Diverged	Noncoding	No Repeats, Between *L31RA* and *DDX4* in *E. telfairi* (NW_004558716.1), 100% ID Across 25bp
336666	Diverged	Noncoding	No Repeats, Between *ZBTB17* and *HSPB7* in GG (NC_006108.3), 74% ID Across 416bp[5]
34426	Diverged	Unknown	No Repeats, Unknown Orthology, Best Hit to Zebra (NW_004531880.1), 93% ID Across 30bp
344738	Diverged	Unknown	No Repeats, Unknown Orthology, Short Hit to *M. Ochrogaster* Unplaced Scaffold (NW_004949166.1), 93% ID Across 30bp
346288[7]	Diverged	*BMPER*	(I) No Repeats, Unknown Orthology, Short Hit to *M. Musculus* Strain C57BL/6J (NT_039472.8), 89% ID Across 36bp
351675	Diverged	Noncoding	No Repeats, Unknown Orthology
357822	Diverged	Noncoding	No Repeats. Unknown Orthology, Short Hit to *P. anubis* (NW_003877394.1), 91% ID Across 35bp

357734	Diverged	Noncoding	No Repeats, Between Forkhead Box Protein L1-like (LOC100288524) and *ZDHHC7* in *X. tropicalis* (NW_004668236.1), 79%ID Across 52bp
363903	Diverged	Noncoding	No Repeats, Flanked by *GA27785* on *Drosophila pseudoobscura* strain MV2-25 (NW_001589783.2), 80% ID Across 60bp
370083	Diverged	Unknown	No Repeats, Unknown Orthology, Short Hit to Walrus (NW_004450896.1) , 93% ID Across 30bp
370977	Diverged	Noncoding	No Repeats, Between *CDH2* and *DSC2* in GG (NC_006089.3), 84%ID Across 301bp[5]
54547	Diverged	Noncoding	High Repeats, Turkey_2.01 (NW_003435376.1), 77% ID Across 373bp
63584	Diverged	Unknown	High Repeats, Best Hit to GG (XM_423233.4), 81% ID Across 502bp
85406	Diverged	Noncoding	No Repeats, Unknown Orthology, Best Hit to *X. tropicalis* (NW_004675458.1), 72% ID Across 169bp
93354	Diverged	Noncoding	High Interspersed Repeats, Between LOC101749928 and Envelope Glycoprotein Gp95-like in GG (LOC101750146), NC_006088.3, 85% ID Across 525bp
254520	Diverged	Noncoding	No Repeats, Between *ZNF706* (Loc101560117) and *TGIF2LX* (LOC101589443) in *O.degus*, NW_004524679.1, 91% ID Across 31bp
332175	Diverged	Noncoding	No Repeats, Unknown Orthology, Short Hit to *M. rotundata* (NW_003797141.1), 98% ID Across 35bp

[1] NB1.0 simple *de novo* contig ID.
[2] The direction of the outlier in the full blastn distribution for the comparative genome alignment with zebra finch.
[3] Concise prediction (top blastn hit) of the genomic information content for each contig (gene symbol, noncoding, or unknown).
[4] Detailed description of the genomic information content for each contig, as evidenced by blastn searches of refseq_genomic, reseq_rna, and nr/nt, with repeat content predicted by RepeatMasker. Outliers for conservation were annotated based on the Zebra Finch (ZF) Genome. GG indicates *Gallus gallus*. (I) indicates intron(s), (E) indicates exon(s), and (I, E) indicates both. Note, the blast databases are dynamic, and therefore, descriptions correspond to results achieved at the time of analysis (Chicken 4.0 and Zebra Finch Build 3.2.4).
[5] Genes are predicted to be syntenic and proximal in both the chicken and zebra finch genomes via blastn and/or NCBI Map Viewer.
[6] Synteny and proximity of genes could not be conclusively determined using the chicken and zebra finch genome resources.
[7] Used corresponding scaffolds to confirm the predicted intron sequence.
[8] Gene/genomic region was found to also be an outlier in the scarlet macaw genome analysis (Seabury et al. 2013).
[9] Previously reported to be under purifying selection (see references).

References for Loci Previously Reported To Be Under Purifying Selection

Bakewell MA. Genomic Patterns of Gene Evolution. Doctoral dissertation. Albuquerque: University of New Mexico; 2011. [***ALPK3***]

Benn M. Apolipoprotein B levels, *APOB* alleles, and risk of ischemic cardiovascular disease in the general population, a review. Atherosclerosis. 2011;206:17-30. [***APOB***]

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, et al. Natural selection on protein-coding genes in the human genome. Nature. 2005;437:1153-1157. [***LPL***]

Goto H, Watanabe K, Araragi N, Kageyama R, Tanaka K, et al. The identification and functional implications of human-specific. BMC Evolutionary Biology. 2009;9:224. [***GRM3***]

Hett AK, Pitra C, Jenneckens I, Ludwig A. Characterization of *SOX9* in European Atlantic sturgeon (Acipenser sturio). Journal of Heredity. 2005;96:150-154. [***SOX9***]

Irimia M, Maeso I, Burguera D, Hidalgo-Sánchez M, Puelles L, et al. Contrasting 5' and 3'evolutionary histories and frequent evolutionary convergence in Meis/hth gene structures. Genome Biology and Evolution; 2011;3:551. [***MEIS2***]

Lawson HA. Molecular Evolutionary Underpinnings of Craniofacial Growth and Development. Ann Arbor: ProQuest; 2008. [***CFDP1***]

Lin MF, Kheradpour P, Washietl S, Parker BJ, Pedersen JS, et al. Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes. Genome Research. 2011;21:1916-1928. [***CADPS***]

Mikheyev AS, Mueller UG, Abbot P. Cryptic sex and many-to-one coevolution in the fungus-growing ant symbiosis. Proceedings of the National Academy of Sciences. 2006;103:10702-10706. [***RAD51B***]

Paixão-Côrtes VR, Salzano FM, Bortolini MC. Evolutionary History of Chordate *PAX* Genes: Dynamics of Change in a Complex Gene Family. PloS ONE. 2013;8:e73560. doi: 10.1371/journal.pone.0073560. [***PAX2***]

Schleinitz D, Klöting N, Böttcher Y, Wolf S, Dietrich K, et al. Genetic and evolutionary analyses of the human bone morphogenetic protein receptor 2 (*BMPR2*) in the pathophysiology of obesity. PloS ONE. 2011;6:e16155. doi: 10.1371/journal.pone.0016155. [***BMPR2***]

Sharp PM, Li WH. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Molecular Biology and Evolution. 1987;43:222-230. [***LPP***]

Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, et al. Strong purifying selection in transmission of mammalian mitochondrial DNA. PLoS Biology. 2008;6:e10. doi: 10.1371/journal.pbio.0060010. [**Mitochondria**]

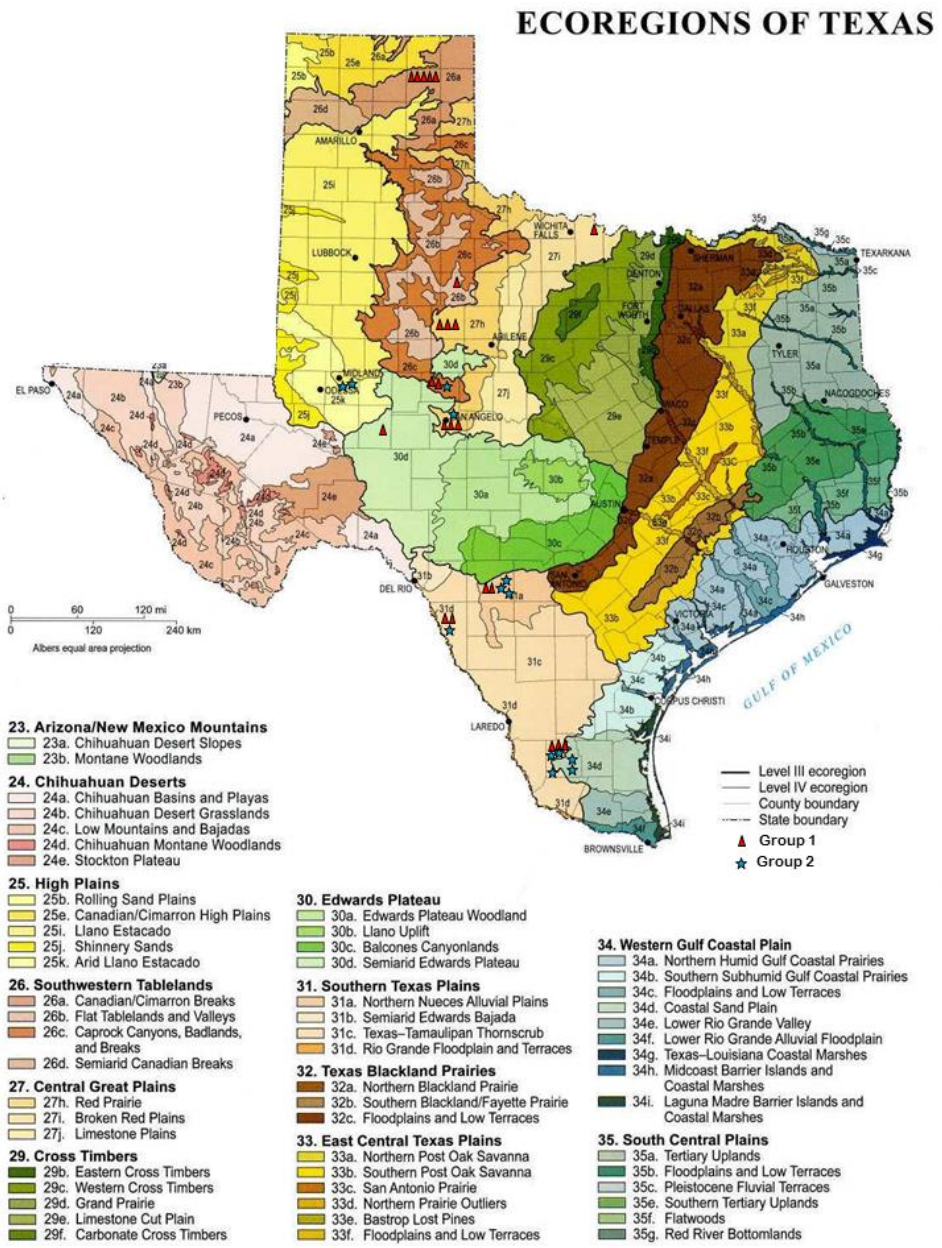
Sugiura N, Dadashev V, Corriveau RA. *NARG2* encodes a novel nuclear protein with (S/T) PXX motifs that is expressed during development. European Journal of Biochemistry. 2004;271:4629-4637. [***NARG2***]

Tabata H, Hachiya T, Nagata KI, Sakakibara Y, Nakajima K. Screening for candidate genes involved in the expansion of the cerebral cortex during evolution by combining expression and evolutionary analyses. Frontiers in Neuroanatomy. 2013;doi: 10.3389/fnana.2013.00024. [***JAG1***]

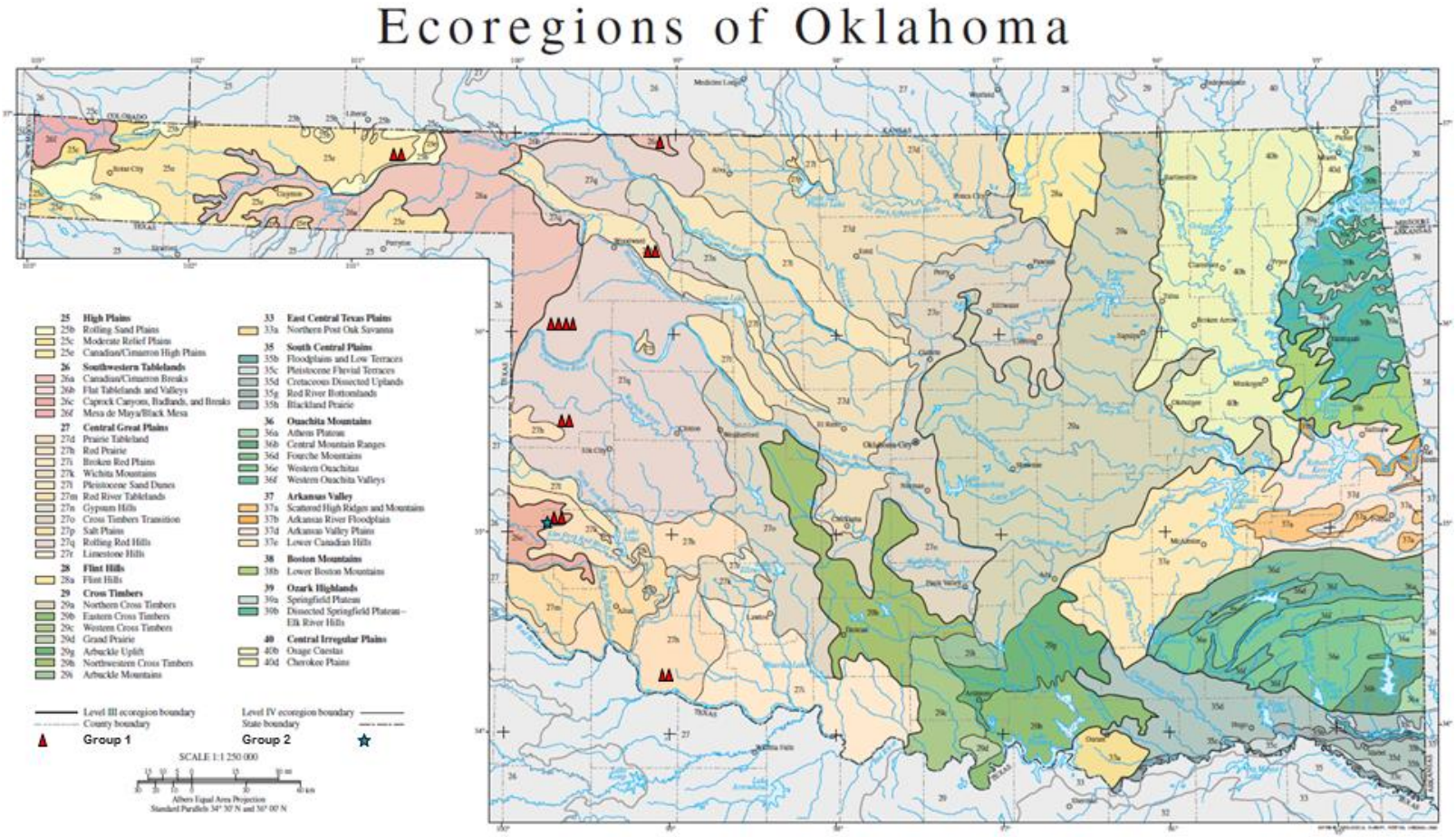
Wang L, Li G, Wang J, Ye S, Jones G, et al. Molecular cloning and evolutionary analysis of the *GJA1* (connexin43) gene from bats (Chiroptera). Genetic Research. 2009;91:101. [***GJA1***]

Wu W, de Folter S, Shen X, Zhang W, Tao S. Vertebrate paralogous *MEF2* genes: Origin, conservation, and evolution. PLoS ONE. 2011;6:e17334. doi: 10.1371/journal.pone.0017334. [***MEF2A***; ***MEF2C***]

APPENDIX B



Map of Texas Bobwhite Specimen Localities Based on U.S. Environmental Protection Agency Level III Ecoregions (adapted from http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). Red triangles represent the collection localities for Group 1 birds. Blue stars represent the collection localities for Group 2 birds.



Map of Oklahoma Bobwhite Specimen Localities Based on U.S. Environmental Protection Agency Level III Ecoregions (adapted from http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). Red triangles represent the collection localities for Group 1 birds. Blue stars represent the collection localities for Group 2 birds.

APPENDIX C

Comparative Analyses of Diversity for Partial and Complete Mitogenomes.

Summary Data*	Partial D-Loop	Complete D-Loop	Complete Mitogenome
Sample size (Haplotypes)	53	53	53
Size of analyzed region (bp)	353	1,152	16,702
Total variable Sites	19	31	323
Total number of mutations	20	33	326
Total unique haplotypes	21	32	49
Haplotype diversity (hd)	0.855	0.963	0.997
Nucleotide diversity (π)	0.00858	0.00424	0.00358

* Excluding heteroplasmic minor allele haplotypes and gaps.

Pairwise F_{ST} Values for Partial and Complete Mitogenomes of Bobwhites with Geographic Subspecies Designations.

	Partial D-Loop (353bp)		Complete D-Loop (1,152 bp)		Mitogenome (16,709 bp)	
	C. v. texanus	*C. v. taylori*	*C. v. texanus*	*C. v. taylori*	*C. v. texanus*	*C. v. taylori*
C. v. texanus		< 0.0001		≤ 0.0001		≤ 0.0001
C. v. taylori	0.24654*		0.19451*		0.29450*	

* Significant ($P < 0.05$) F_{ST} values (below diagonal, with standard errors above diagonal); excluding heteroplasmic minor allele haplotypes.

Diversity and Demographic Analyses for Complete Mitogenomes of Bobwhites with Geographic Subspecies Designations, as Compared to Classification via Mitogenome Divergence.

Subcategories *	**Hd**	***π***	***D***	F_S
Group 1 + Group 2	0.997	0.00358	-0.6017	-8.845[c]
Group 1	0.996	0.00116	-1.8893[a,c]	-15.679[d]
Group 2	0.989	0.00097	-1.3933	-2.534
C. v. texanus	0.997	0.00430	0.3162	-2.258
C. v. taylori	0.994	0.00167	-2.2034[b,d]	-4.822

* Excluding heteroplasmic minor allele haplotypes.

[a] $P < 0.05$ (beta distribution)

[b] $P < 0.01$ (beta distribution)

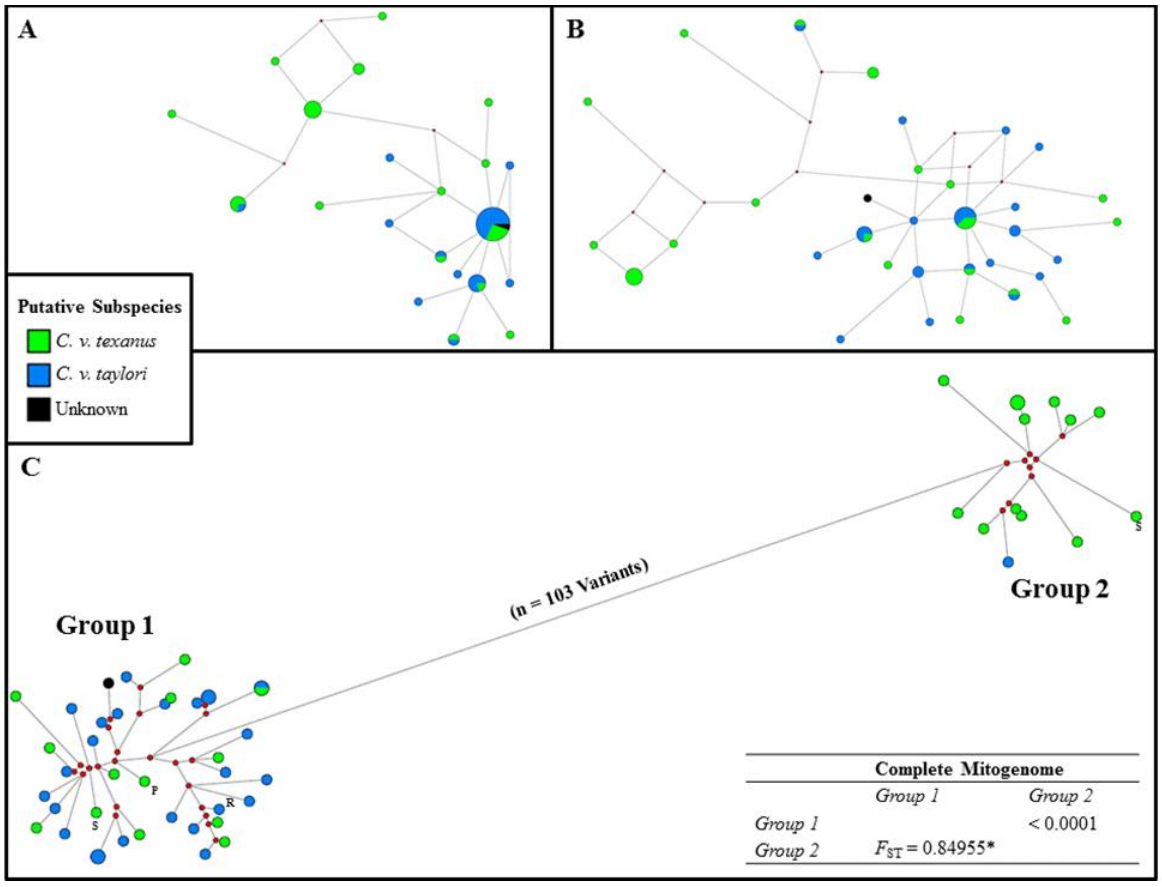
[c] $P < 0.05$ (coalescent simulations)

[d] $P < 0.01$ (coalescent simulations

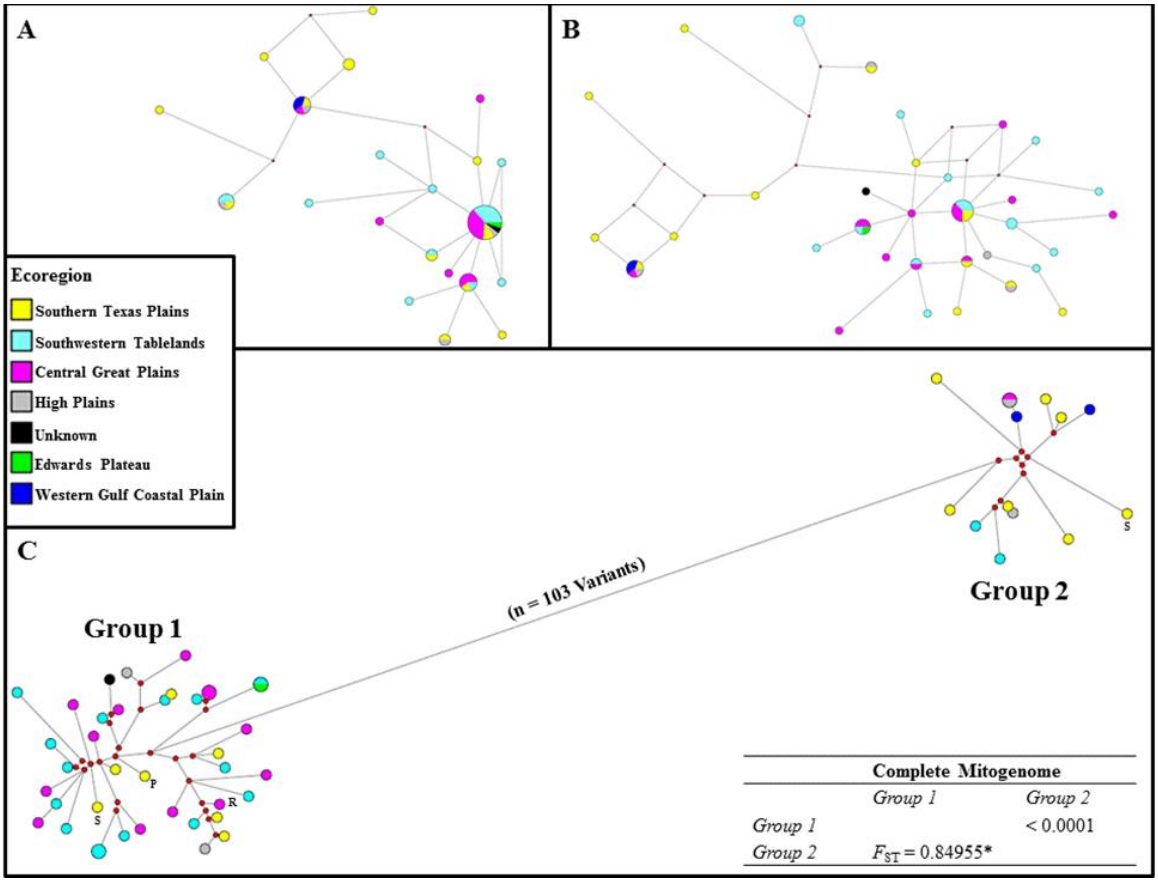
Genetic Differentiation among Divergent Bobwhite Mitogenome Lineages, Excluding Heteroplasmic Minor Allele Haplotypes.

	K_S	K_S*	*Z*	*Z**
Value	18.55985	2.90158	397.34875	5.67334
P	< 0.001	< 0.001	< 0.001	< 0.001

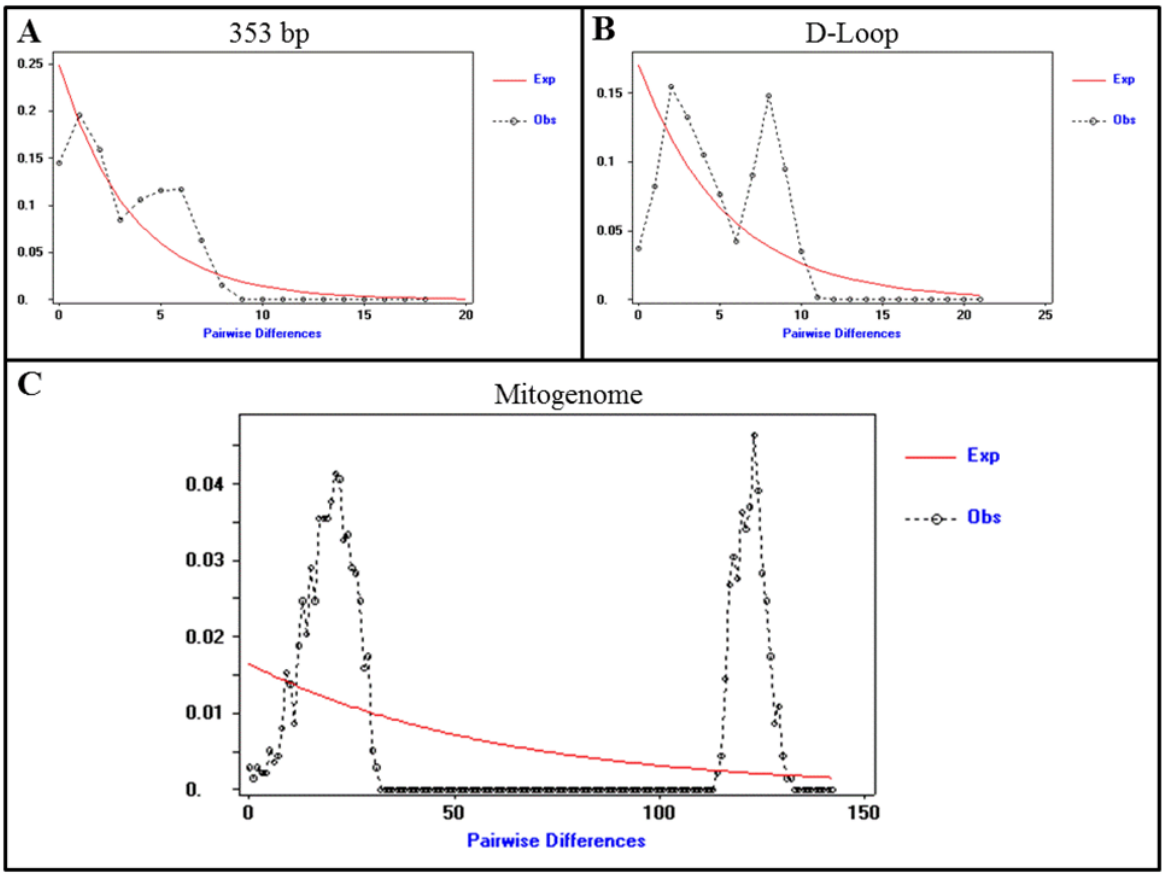
APPENDIX D



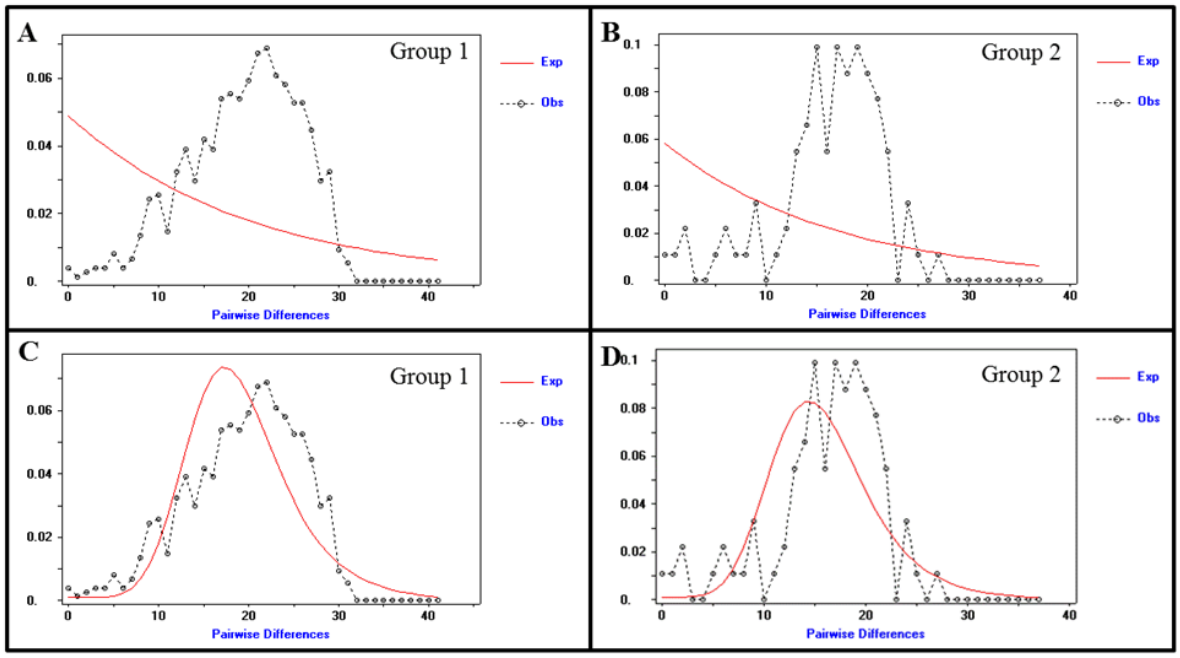
Median Joining (MJ) Haplotype Networks (341) Constructed for Partial and Complete Bobwhite Mitogenome Sequences, Excluding Heteroplasmic Minor Allele Haplotypes, and Color-coded by Geographic Subspecies Designations (5, 9, 10). (A) MJ haplotype network for 353 bp of the mitochondrial D-Loop (9). (B) MJ haplotype network for the complete D-Loop (1152 bp). (C) MJ haplotype network for the complete mitogenome (16,709 bp including gaps). Default weights for SNPs and indels were used (10 and 20, respectively), with node sizes proportional to haplotype frequency, and branch lengths drawn to scale. Red dots indicate median vectors. The complete mitogenome haplotypes were observed to form two divergent clusters (i.e., Group 1, Group 2; n = 103 variants). Pairwise F_{ST} values (below diagonal) with standard errors (above diagonal) were computed to assess genetic differentiation between the two clusters, with the asterisk (*) indicating a significant F_{ST} value ($P < 0.05$). Panel C includes three complete mitogenome haplotypes for bobwhites lawfully harvested from active surrogating pastures (i.e., pen release sites = S), and one haplotype from a lawfully harvested pen-released bobwhite (P). (R) Indicates the reference mitogenome (313).



Median Joining (MJ) Haplotype Networks (341) Constructed for Partial and Complete Bobwhite Mitogenome Sequences, Excluding Heteroplasmic Minor Allele Haplotypes, and Color-coded by the U.S. Environmental Protection Agency Level III Ecoregions (http://www.epa.gov/wed/pages/ecoregions/level_iii_iv.htm). (A) MJ haplotype network for 353 bp of the mitochondrial D-Loop (9) (B) MJ haplotype network for the complete D-Loop (1152 bp). (C) MJ haplotype network for the complete mitogenome (16,709 bp including gaps). Default weights for SNPs and indels were used (10 and 20, respectively), with node sizes proportional to haplotype frequency, and branch lengths drawn to scale. Red dots indicate median vectors. The complete mitogenome haplotypes were observed to form two divergent clusters (i.e., Group 1, Group 2; n = 103 variants). Pairwise F_{ST} values (below diagonal) with standard errors (above diagonal) were computed to assess genetic differentiation between the two clusters, with the asterisk (*) indicating a significant F_{ST} value ($P < 0.05$). Panel C includes three complete mitogenome haplotypes for bobwhites harvested from active surrogating pastures (i.e., pen release sites = S), and one haplotype from a lawfully harvested pen-released bobwhite (P). (R) Indicates the reference mitogenome (313).



Mismatch Distributions for Partial and Complete Bobwhite Mitogenome Sequences, Excluding Heteroplasmic Minor Allele Haplotypes. (A) 353 bp of the mitochondrial D-Loop (9). (B) Complete D-Loop (1,151 bp excluding gaps). (C) Complete mitogenome (16,698 bp excluding gaps). The x-axis represents the number of pairwise differences (mismatches) and the y-axis represents the frequency of these differences. The observed mismatch distribution (dashed line) is compared to the expected distribution (red line) for a stable population (i.e., constant population size).



Bobwhite Historical Demography, as Inferred via Mismatch Distribution for Constant Population Size and Growth-decline Models, Excluding Heteroplasmic Minor Allele Haplotypes. (A) The observed mismatch distribution (dashed line) for bobwhite Group 1 (n = 39) as compared to the expected distribution (red line) for a stable population (i.e., constant population size). (B) The observed mismatch distribution (dashed line) for bobwhite Group 2 (n = 14) as compared to the expected distribution (red line) for a stable population (i.e., constant population size). (C) The observed mismatch distribution (dashed line) for bobwhite Group 1 (n = 39) as compared to the expected distribution (red line) for a growth-decline model. (D) The observed mismatch distribution (dashed line) for bobwhite Group 2 (n = 14) as compared to the expected distribution (red line) for a growth-decline model.